



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. En un grafo una comunidad es:
 - a) Un conjunto de nodos altamente conectados que forman un subgrafo completamente independiente del resto de la red.
 - b) Un conjunto de nodos aislados.
 - c) Un conjunto de nodos que no comparten ninguna similitud estructural o funcional entre sí.
 - d) Un conjunto de nodos que están más densamente interconectados entre sí que con los nodos fuera del conjunto.
2. Sobre WordNet y su aplicación en el procesamiento del lenguaje natural, se puede afirmar que:
 - a) Aunque WordNet es una herramienta valiosa en el NLP, su estructura no incluye información sobre la frecuencia de uso de las palabras en el lenguaje natural.
 - b) El diseño de WordNet facilita su integración en aplicaciones multilingües de NLP, aunque su desarrollo original se centró en el inglés.
 - c) Los synsets facilitan la identificación de relaciones semánticas entre palabras, como la hiperonimia y la meronimia, enriqueciendo tareas de NLP.
 - d) WordNet proporciona una base para la desambiguación semántica al agrupar palabras con significados similares en synsets.
3. En un sistema de control de tráfico urbano basado en reglas, ¿cuál de las siguientes reglas sería más efectiva para manejar situaciones de congestión en una intersección?
 - a) Si hay pocos vehículos en la intersección, reducir el tiempo de los semáforos en verde.
 - b) Si hay muchos vehículos en la intersección, aumentar el tiempo de los semáforos en verde.
 - c) Si hay un vehículo de emergencia en la intersección, detener todos los demás vehículos.
 - d) Si hay muchos peatones cruzando la intersección, reducir el tiempo de los semáforos en rojo.
4. Al realizar la optimización de contenido para SEO debe considerarse:
 - a) Utilizar etiquetas de título y meta descripciones únicas y relevantes para cada página.
 - b) Seleccionar las palabras clave al azar.
 - c) Crear contenido valioso y original que satisfaga las necesidades de los usuarios.
 - d) Incluir palabras clave de manera excesiva en el contenido para mejorar el posicionamiento.
5. ¿Qué es un 'Uniform Resource Locator (URL) Frontier' en el contexto de Web Crawling?
 - a) Una base de datos que almacena URLs únicas identificadas como recursos en la web.
 - b) Una técnica para filtrar URLs irrelevantes y mejorar la eficiencia del crawling.
 - c) Un protocolo que define cómo se deben formatear las URLs para el crawling.
 - d) Una lista prioritaria de URLs que aún no han sido visitadas por el crawler.
6. No se considera como técnica para detectar comunidades en una red:
 - a) Analizar la mutualidad de los enlaces.
 - b) Usar el agrupamiento jerárquico.
 - c) Utilizar el algoritmo de K-Means.
 - d) Encontrar cliques de vértices de grado par.
7. Existen varias estrategias consideradas como no recomendables para mejorar el SEO de un sitio web. Dentro de estas estrategias negativas, se encuentran:
 - a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Evitar el uso de las meta etiquetas y descripciones del sitio web para reflejar el contenido de la página.
 - c) Construir enlaces naturales de sitios web con autoridad y relevancia temática.
 - d) Incluir una cantidad excesiva de palabras clave irrelevantes para intentar manipular los rankings de búsqueda.
8. El componente responsable de la gestión de recursos y planificación de tareas en Hadoop es:
 - a) MapReduce.

- b) HDFS.
 - c) Hadoop Common.
 - d) YARN.
9. El algoritmo Hypertext Induced Topic Selection (HITS) intenta buscar nodos especiales. Estos son conocidos como:
- a) Sensibles.
 - b) Centrales.
 - c) Hubs.
 - d) Autoridades.
10. ¿Qué algoritmos permiten obtener información de una red?
- a) Índices de centralidad.
 - b) Detección de comunidades.
 - c) Hypertext Induced Topic Selection (HITS).
 - d) PageRank.
11. ¿Cuál de las siguientes opciones describe mejor la diferencia clave entre Web Crawling y Web Scraping?
- a) Web Scraping implica el análisis de la estructura y el contenido de las páginas web para extraer datos, mientras que Web Crawling se refiere a la descarga y almacenamiento de páginas web completas.
 - b) Web Scraping es más eficaz para rastrear e indexar contenido web para motores de búsqueda, mientras que Web Crawling se utiliza principalmente para la extracción de datos en proyectos de investigación.
 - c) Web Crawling se centra en la exploración y recopilación de enlaces de múltiples sitios web, mientras que Web Scraping se enfoca en la extracción específica de datos de páginas web individuales.
 - d) Web Crawling se realiza utilizando herramientas de automatización como Selenium WebDriver, mientras que Web Scraping se lleva a cabo mediante el análisis de HTML y CSS.
12. Se tiene una red de ingredientes donde cada uno representa un nodo y las aristas simbolizan que los ingredientes forman parte de una misma receta. Se busca mejorar la experiencia culinaria mediante la elaboración de combinaciones de ingredientes más interesantes y creativas, para lo cual se debe:
- a) No tener en cuenta la centralidad de grado de los ingredientes para no darle mayor importancia a los ingredientes más comunes.
 - b) No considerar la centralidad de grado de los ingredientes, ya que todas las combinaciones de ingredientes son igualmente válidas.
 - c) Utilizar la centralidad de grado para identificar los ingredientes más populares en la red y crear combinaciones que incluyan una variedad de ingredientes menos comunes.
 - d) Utilizar la centralidad de grado para identificar los ingredientes menos conectados en la red y tomarlos en cuenta para su inclusión en las combinaciones.
13. Considerando las dimensiones y desafíos inherentes a Big Data puede afirmarse, tomando en cuenta las características clave y las implicaciones para su procesamiento y análisis, que:
- a) Big Data se caracteriza principalmente por su pequeño volumen y uniformidad, permitiendo un procesamiento eficiente con mínimas adaptaciones de las herramientas de análisis de datos tradicionales.
 - b) Big Data no desafía la capacidad de las herramientas tradicionales de procesamiento de datos para capturar, almacenar, gestionar y analizar efectivamente la información, dada la evolución constante de las capacidades computacionales y algoritmos de optimización.
 - c) Aunque Big Data puede incluir datos estructurados, su naturaleza se expande al incorporar grandes cantidades de datos no estructurados y semiestructurados, lo que exige el uso de tecnologías especializadas en almacenamiento y procesamiento como Hadoop y sistemas de bases de datos NoSQL.
 - d) Además de su complejidad y diversidad, Big Data introduce desafíos significativos en términos de veracidad y variabilidad al requerir métodos avanzados de limpieza y validación de datos para asegurar la integridad del análisis.
14. Una plataforma de comercio electrónico desea mejorar la experiencia del usuario al permitir una navegación más personalizada y contextualizada. Actualmente los usuarios tienen dificultades para encontrar productos relevantes debido a la gran cantidad de opciones disponibles. La empresa está interesada en implementar características de la Web 2.5 y la Web Semántica para abordar este problema. ¿Qué características podrían ayudar para ofrecer una navegación más personalizada y contextualizada?
- a) La implementación de ontologías y metadatos para enriquecer la descripción de productos y mejorar la precisión de las recomendaciones.
 - b) La optimización de la velocidad de carga del sitio web para mejorar la experiencia del usuario y reducir el abandono del carrito de compra.
 - c) La incorporación de los productos en tendencia en el mercado.

- d) La integración de redes sociales para permitir la recomendación de productos basada en las preferencias de amigos y contactos.
15. En el contexto de la representación del conocimiento basada en herencia, ¿qué caracteriza a la herencia cancelable?
- a) Las propiedades heredadas siempre se mantienen y no pueden anularse.
 - b) La herencia es el resultado del razonamiento no transitivo.
 - c) Las conclusiones no están determinadas y dependen del nodo de interés.
 - d) No existe ambigüedad en las conclusiones obtenidas.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Se puede afirmar que:
 - a) Uno de los desafíos en el procesamiento de Big Data es la capacidad de gestionar y analizar datos provenientes de diversas fuentes y en diferentes formatos de manera eficiente.
 - b) MapReduce es un modelo de procesamiento distribuido utilizado para trabajar con grandes volúmenes de datos.
 - c) Los SRI pueden beneficiarse de MapReduce para mejorar la RI relevante.
 - d) El término “Big Data” se refiere exclusivamente al volumen de datos que una organización maneja, sin tener en cuenta la velocidad, la variedad y la veracidad de los datos.
2. La Web 1.0 se caracteriza por:
 - a) Los propietarios de los sitios web proporcionan contenido de forma periódica.
 - b) Uso de comunidades virtuales para popularizar los sitios web de noticias.
 - c) Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
 - d) Los sitios web se centran en brindar información en lugar de facilitar la colaboración o participación de los usuarios.
3. Si se tiene el conjunto de páginas interconectadas $\{A \rightarrow B, C; B \rightarrow C; C \rightarrow A\}$, entonces la página con valor más alto de PageRank es:
 - a) B.
 - b) C.
 - c) A y B.
 - d) A.
4. ¿Qué algoritmos permiten obtener información de una red?
 - a) Índices de centralidad.
 - b) Detección de comunidades.
 - c) Hypertext Induced Topic Selection (HITS).
 - d) PageRank.
5. Considerando las dimensiones y desafíos inherentes a Big Data puede afirmarse, tomando en cuenta las características clave y las implicaciones para su procesamiento y análisis, que:
 - a) Big Data se caracteriza principalmente por su pequeño volumen y uniformidad, permitiendo un procesamiento eficiente con mínimas adaptaciones de las herramientas de análisis de datos tradicionales.
 - b) Big Data no desafía la capacidad de las herramientas tradicionales de procesamiento de datos para capturar, almacenar, gestionar y analizar efectivamente la información, dada la evolución constante de las capacidades computacionales y algoritmos de optimización.
 - c) Aunque Big Data puede incluir datos estructurados, su naturaleza se expande al incorporar grandes cantidades de datos no estructurados y semiestructurados, lo que exige el uso de tecnologías especializadas en almacenamiento y procesamiento como Hadoop y sistemas de bases de datos NoSQL.
 - d) Además de su complejidad y diversidad, Big Data introduce desafíos significativos en términos de veracidad y variabilidad al requerir métodos avanzados de limpieza y validación de datos para asegurar la integridad del análisis.
6. La afirmación que mejor refleja el principio subyacente de PageRank, considerando su importancia en la clasificación de los sitios web, es:
 - a) El algoritmo de PageRank considera tanto la cantidad como la calidad de los enlaces entrantes, asignando mayor valor a los enlaces provenientes de sitios web considerados como “importantes”.
 - b) PageRank valora más la cantidad de enlaces entrantes a una página web, independientemente de la calidad o relevancia de estos enlaces.
 - c) La efectividad de PageRank se basa exclusivamente en el análisis de las palabras clave contenidas en los enlaces entrantes, sin considerar la estructura de enlace de la Web.
 - d) PageRank opera bajo el supuesto de que los enlaces entrantes y salientes tienen el mismo im-

- pacto en la valoración de la relevancia de una página web.
7. En el contexto de la representación del conocimiento basada en herencia, ¿qué caracteriza a la herencia cancelable?
 - a) Las propiedades heredadas siempre se mantienen y no pueden anularse.
 - b) La herencia es el resultado del razonamiento no transitivo.
 - c) Las conclusiones no están determinadas y dependen del nodo de interés.
 - d) No existe ambigüedad en las conclusiones obtenidas.
 8. Un SRI es capaz de:
 - a) No necesitar de ningún almacenamiento externo para alojar los índices de los datos.
 - b) Reducir el tiempo de indexación de los datos si utiliza servidores distribuidos dentro de la red para que cada uno ejecute la indexación del mismo conjunto de datos.
 - c) Generar índices invertidos de manera óptima sin considerar el contexto.
 - d) Crear los índices asociados a los datos sin tener que analizar cada dato.
 9. El algoritmo de PageRank converge si:
 - a) Se define un factor de normalización en la fórmula de la función.
 - b) La norma de la diferencia entre los vectores es menor a un umbral predefinido.
 - c) Finaliza la ejecución cuando el número de iteraciones excede un máximo de iteraciones previamente definido.
 - d) El algoritmo no se implementa de forma iterativa.
 10. En una empresa comercial, se tiene una red donde los nodos corresponden a los empleados y las aristas representan la frecuencia con la que colaboran en las ventas. Se busca mejorar la colaboración entre los empleados para aumentar las ventas totales, por lo que la directiva debe:
 - a) No tener en cuenta la centralidad de cercanía de los empleados al diseñar estrategias de colaboración, ya que esto podría introducir complicaciones adicionales en el proceso de trabajo.
 - b) Utilizar la centralidad de cercanía para identificar los empleados menos cercanos a otros en la red y asignarles tareas individuales para evitar posibles conflictos y desacuerdos en el proceso de colaboración.
 - c) No considerar la centralidad de cercanía de los empleados al diseñar estrategias de colaboración, ya que todos los empleados tienen el mismo acceso a la información y los recursos.
 - d) Utilizar la centralidad de cercanía para identificar los empleados más cercanos a otros en la red y promover la colaboración entre ellos, facilitando así la comunicación y el intercambio de conocimientos para mejorar las ventas.
 11. La web actual se enfrenta a problemas como:
 - a) Presencia de una alta calidad en los datos.
 - b) Heterogeneidad en los datos.
 - c) Presencia de datos volátiles y distribuidos.
 - d) Presencia de grandes volúmenes de datos estructurados.
 12. En una red de transporte donde cada nodo es una parada de autobús y las aristas representan si existe un carro que pasa por ambos sitios, ¿qué puede mejorar el sistema de transporte?
 - a) Utilizar la centralidad de intermediación para identificar las paradas de transferencia clave y establecer nuevas conexiones entre diferentes líneas de transporte público.
 - b) Utilizar el grafo inducido de los nodos con mayor valor en la centralidad de grado para aplicar la centralidad de intermediación con el propósito de reforzar las paradas con mayor tráfico.
 - c) No implementar sistemas de información en tiempo real para los usuarios, ya que pueden aumentar la carga de los trabajadores.
 - d) No considerar la centralidad de intermediación de las estaciones de transporte público al planificar rutas y horarios, ya que no tiene impacto en las conexiones entre las paradas.
 13. Un sistema cuenta con la siguiente información:
 - Juan pertenece a la marina.
 - Juan es capellán (sacerdote encargado de una tarea específica fuera de la parroquia).
 - Los infantes de la marina suelen ser bebedores de cerveza.
 - Un capellán no suele ser bebedor de cerveza.
 - Un bebedor de cerveza suele tener sobrepeso.
 - Por lo general, un infante de la marina está en buena condición física.

Si se representa la información en una red de herencia se puede concluir que:

 - a) Existen dos razonamientos cancelables.
 - b) Juan tiene sobrepeso.
 - c) No se puede asegurar que Juan sea capellán.
 - d) No se puede asegurar que Juan esté en buena condición física.

14. No se considera como técnica para detectar comunidades en una red:
 - a) Analizar la mutualidad de los enlaces.
 - b) Usar el agrupamiento jerárquico.
 - c) Utilizar el algoritmo de K-Means.
 - d) Encontrar cliques de vértices de grado par.
15. Una buena práctica de SEO para mejorar el posicionamiento de un sitio web en los motores de búsqueda es:
 - a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Copiar contenido directamente de otros sitios web populares para aumentar la cantidad de páginas indexadas.
 - c) Llenar el contenido con palabras clave irrelevantes para aumentar la densidad de palabras clave.
 - d) Obtener enlaces de otros sitios web relevantes y de calidad que apunten al sitio.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. En la RI en el contexto de Big Data, se puede asegurar que:
 - a) El almacenamiento distribuido es una técnica obsoleta.
 - b) El almacenamiento distribuido centraliza todos los datos en un único servidor para facilitar su gestión y mantenimiento.
 - c) El almacenamiento distribuido reparte los datos en múltiples servidores para mejorar la disponibilidad y la redundancia del sistema.
 - d) El almacenamiento distribuido ofrece ventajas significativas en términos de escalabilidad y rendimiento en comparación con el almacenamiento centralizado.
2. Una plataforma de comercio electrónico desea mejorar la experiencia del usuario al permitir una navegación más personalizada y contextualizada. Actualmente los usuarios tienen dificultades para encontrar productos relevantes debido a la gran cantidad de opciones disponibles. La empresa está interesada en implementar características de la Web 2.5 y la Web Semántica para abordar este problema. ¿Qué características podrían ayudar para ofrecer una navegación más personalizada y contextualizada?
 - a) La implementación de ontologías y metadatos para enriquecer la descripción de productos y mejorar la precisión de las recomendaciones.
 - b) La optimización de la velocidad de carga del sitio web para mejorar la experiencia del usuario y reducir el abandono del carrito de compra.
 - c) La incorporación de los productos en tendencia en el mercado.
 - d) La integración de redes sociales para permitir la recomendación de productos basada en las preferencias de amigos y contactos.
3. Los Web Crawlers se enfrentan a desafíos constantes. Dentro de ellos se encuentran:
 - a) La falta de acceso a la base de datos del servidor web para extraer información actualizada.
 - b) La modificación del código y la estructura del sitio web.
 - c) La incapacidad para interpretar correctamente el lenguaje de programación utilizado en el desarrollo de los sitios web.
 - d) La dificultad para generar el contenido dinámico en tiempo real.
4. El uso de Hadoop y MapReduce en el contexto de la RI tiene como objetivo:
 - a) Facilitar el análisis en tiempo real de datos de redes sociales.
 - b) Automatizar el mantenimiento de sistemas de bases de datos.
 - c) Mejorar la eficiencia energética en centros de datos.
 - d) Procesar y analizar grandes conjuntos de datos para la RI.
5. No se considera como técnica para detectar comunidades en una red:
 - a) Analizar la mutualidad de los enlaces.
 - b) Usar el agrupamiento jerárquico.
 - c) Utilizar el algoritmo de K-Means.
 - d) Encontrar cliques de vértices de grado par.
6. En la RI en el contexto de Big Data, se puede asegurar que:
 - a) La indexación distribuida es una técnica obsoleta.
 - b) La indexación distribuida almacena todos los datos en un solo servidor para facilitar su acceso y búsqueda.
 - c) La indexación distribuida divide los datos en múltiples fragmentos que se almacenan en varios nodos para permitir búsquedas paralelas y mejorar la escalabilidad.
 - d) La indexación distribuida no ofrece ventajas en términos de rendimiento y escalabilidad en comparación con la indexación centralizada.
7. Una buena práctica de SEO para mejorar el posicionamiento de un sitio web en los motores de búsqueda es:
 - a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.

- b) Copiar contenido directamente de otros sitios web populares para aumentar la cantidad de páginas indexadas.
 - c) Llenar el contenido con palabras clave irrelevantes para aumentar la densidad de palabras clave.
 - d) Obtener enlaces de otros sitios web relevantes y de calidad que apunten al sitio.
8. Referente al Web Crawler puede afirmarse que:
- a) Los hipervínculos encontrados en cada sitio web que no pertenecen al dominio donde fueron encontrados se desechan, puesto que no expande el conjunto de URLs sin visitar.
 - b) Las páginas visitadas no se procesan nunca más.
 - c) No necesita de un conjunto inicial de URLs para recorrer la Web.
 - d) No tiene como objetivo indexar y recopilar información de diferentes sitios web.
9. Existen varias estrategias consideradas como no recomendables para mejorar el SEO de un sitio web. Dentro de estas estrategias negativas, se encuentran:
- a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Evitar el uso de las meta etiquetas y descripciones del sitio web para reflejar el contenido de la página.
 - c) Construir enlaces naturales de sitios web con autoridad y relevancia temática.
 - d) Incluir una cantidad excesiva de palabras clave irrelevantes para intentar manipular los rankings de búsqueda.
10. La centralidad de intermediación de un nodo indica:
- a) La frecuencia con la que un nodo actúa como puente en el camino más corto entre otros dos nodos.
 - b) El número total de conexiones entrantes y salientes.
 - c) La resistencia del nodo a fallos.
 - d) La cantidad de nodos vecinos directos.
11. Un sistema cuenta con la siguiente información:
- Juan pertenece a la marina.
 - Juan es capellán (sacerdote encargado de una tarea específica fuera de la parroquia).
 - Los infantes de la marina suelen ser bebedores de cerveza.
 - Un capellán no suele ser bebedor de cerveza.
 - Un bebedor de cerveza suele tener sobrepeso.
 - Por lo general, un infante de la marina está en buena condición física.
- Si se representa la información en una red de herencia se puede concluir que:
- a) Existen dos razonamientos cancelables.
 - b) Juan tiene sobrepeso.
 - c) No se puede asegurar que Juan sea capellán.
 - d) No se puede asegurar que Juan esté en buena condición física.
12. Con respecto a WordNet y su contribución a las ontologías y la representación del conocimiento en los sistemas de RI, puede afirmarse que:
- a) En WordNet, una palabra está asociada a un conjunto de sinónimos (synsets), siendo estas palabras intercambiables en un contexto.
 - b) WordNet diferencia claramente entre los significados de palabras según su uso en diferentes contextos, lo que permite aplicaciones avanzadas en desambiguación semántica más allá de los sistemas de recomendación y RI.
 - c) La integración de WordNet en sistemas de RI permite la expansión de consultas y la mejora de la precisión de los resultados al entender mejor el significado de los términos de búsqueda a través de su contexto semántico y las relaciones entre palabras.
 - d) Una limitación de WordNet en la representación del conocimiento es su enfoque en el idioma inglés, lo que plantea desafíos en la aplicación global y la interoperabilidad con sistemas de información multilingües.
13. La Web 1.0 se caracteriza por:
- a) Los propietarios de los sitios web proporcionan contenido de forma periódica.
 - b) Uso de comunidades virtuales para popularizar los sitios web de noticias.
 - c) Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
 - d) Los sitios web se centran en brindar información en lugar de facilitar la colaboración o participación de los usuarios.
14. En una red de transporte donde cada nodo es una parada de autobús y las aristas representan si existe un carro que pasa por ambos sitios, ¿qué puede mejorar el sistema de transporte?
- a) Utilizar la centralidad de intermediación para identificar las paradas de transferencia clave y establecer nuevas conexiones entre diferentes líneas de transporte público.

- b) Utilizar el grafo inducido de los nodos con mayor valor en la centralidad de grado para aplicar la centralidad de intermediación con el propósito de reforzar las paradas con mayor tráfico.
 - c) No implementar sistemas de información en tiempo real para los usuarios, ya que pueden aumentar la carga de los trabajadores.
 - d) No considerar la centralidad de intermediación de las estaciones de transporte público al planificar rutas y horarios, ya que no tiene impacto en las conexiones entre las paradas.
15. Considerando las prácticas éticas y legales en el Web Scraping, se puede asegurar que:
- a) Web Scraping sobre datos personales sin consentimiento es generalmente aceptado si los datos se utilizan con fines de investigación.
 - b) Es importante revisar y respetar los términos de servicio del sitio web, así como las leyes aplicables de protección de datos y derechos de autor, antes de realizar Web Scraping.
 - c) Web Scraping requiere considerar las políticas de `robots.txt` del sitio web objetivo.
 - d) La extracción de datos mediante Web Scraping siempre es legal, independientemente de las leyes locales sobre derechos de autor y privacidad de datos definidos en los sitios web.

Trabajo de Control
Sistemas de Recuperación de Información
Fecha: 1 de abril de 2024
Temario #4



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. La Web 3 se conoce como:
 - a) Web de solo lectura.
 - b) Internet de las cosas.
 - c) Web de escritura-lectura.
 - d) Web semántica.
2. En una biblioteca digital se necesita implementar un sistema que permita a los usuarios encontrar libros y artículos científicos de forma rápida y precisa. Los documentos están en diversos formatos, incluyendo PDF, EPUB y HTML. Se requiere seleccionar un algoritmo de indexación adecuado para el sistema, por lo que el programador designado para la implementación debe considerar:
 - a) La capacidad para manejar documentos en diferentes formatos.
 - b) La velocidad de indexación y recuperación de datos.
 - c) La complejidad del algoritmo en términos de implementación y mantenimiento.
 - d) La capacidad del algoritmo para procesar imágenes incrustadas o referenciadas en los ficheros.
3. Sobre el algoritmo de PageRank, visto en clase, se puede afirmar que:
 - a) Evalúa la importancia de un sitio web en función de la calidad y cantidad de enlaces entrantes que recibe de otros sitios web.
 - b) Asigna una puntuación baja a los sitios web que contienen muchos enlaces salientes, ya que indica una falta de relevancia.
 - c) Asigna una puntuación alta a los sitios web que tienen un gran número de enlaces entrantes sin tener en cuenta la calidad de esos enlaces.
 - d) Solo tiene en cuenta el contenido en un sitio web para determinar su relevancia en los resultados de búsqueda.
4. Referente al Web Crawler puede afirmarse que:
 - a) Los hipervínculos encontrados en cada sitio web que no pertenecen al dominio donde fueron encontrados se desechan, puesto que no expande el conjunto de URLs sin visitar.
 - b) Las páginas visitadas no se procesan nunca más.
 - c) No necesita de un conjunto inicial de URLs para recorrer la Web.
 - d) No tiene como objetivo indexar y recopilar información de diferentes sitios web.
5. La premisa básica del algoritmo de PageRank para clasificar páginas web en los resultados de búsqueda es:
 - a) Los enlaces entrantes a una página web desde otras páginas contribuyen a su importancia y clasificación.
 - b) La antigüedad de la página web es el principal factor para determinar su clasificación.
 - c) La cantidad de visitas que recibe una página web determina su clasificación.
 - d) El contenido y la relevancia de las palabras clave en la página web determinan su posición.
6. Si una red cumple la propiedad de ser un grafo de mundo pequeño, entonces se conoce que:
 - a) La red posee pocos vértices.
 - b) La longitud media del camino entre todo par de vértices es pequeña.
 - c) El número de componentes fuertemente conexas está relacionado con la cantidad de grafos K_n presentes.
 - d) El grafo es un anillo regular de grado 5.
7. Los algoritmos para detectar comunidades en una red intentan:
 - a) Buscar subgrafos tal que no incluyan nodos cuya ausencia desconecte al subgrafo.
 - b) Seleccionar aleatoriamente nodos de alto grado y sus vecinos.
 - c) Buscar conjuntos donde cada nodo de un mismo conjunto tenga características similares al resto de los nodos del conjunto.
 - d) Encontrar grupos donde los nodos pertenecientes a los mismos grupos son cercanos bajo cierta métrica y lejanos con respecto a los nodos de otros grupos.
8. En el contexto de la RI en redes, se puede afirmar que:

- a) La detección de comunidades ayuda a identificar grupos de nodos altamente conectados entre sí, lo que puede ser útil para comprender la estructura y el contenido de la red.
- b) La detección de comunidades en una red siempre produce resultados objetivos y consistentes independientemente del algoritmo utilizado.
- c) La detección de comunidades solo se aplica a redes pequeñas y simples, no a redes grandes y complejas.
- d) La detección de comunidades no es relevante para la RI en redes, ya que se centra únicamente en la estructura de la red sin considerar el contenido.
9. En un sistema donde el conocimiento está orientado a objetos se puede asegurar que:
- a) Los marcos y las bandas son las estructuras utilizadas para la representación del modelo.
- b) Las funciones de agregación dificultan poder establecer relaciones entre los objetos.
- c) Se enfatiza la atención a la información de la cual se extrajo el conocimiento.
- d) No es posible definir especializaciones de los objetos de la vida real dentro del sistema.
10. ¿Cuál de las siguientes opciones describe mejor la diferencia clave entre Web Crawling y Web Scraping?
- a) Web Scraping implica el análisis de la estructura y el contenido de las páginas web para extraer datos, mientras que Web Crawling se refiere a la descarga y almacenamiento de páginas web completas.
- b) Web Scraping es más eficaz para rastrear e indexar contenido web para motores de búsqueda, mientras que Web Crawling se utiliza principalmente para la extracción de datos en proyectos de investigación.
- c) Web Crawling se centra en la exploración y recopilación de enlaces de múltiples sitios web, mientras que Web Scraping se enfoca en la extracción específica de datos de páginas web individuales.
- d) Web Crawling se realiza utilizando herramientas de automatización como Selenium WebDriver, mientras que Web Scraping se lleva a cabo mediante el análisis de HTML y CSS.
11. Un sistema cuenta con la siguiente información:
- Juan pertenece a la marina.
 - Juan es capellán (sacerdote encargado de una tarea específica fuera de la parroquia).
 - Los infantes de la marina suelen ser bebedores de cerveza.
 - Un capellán no suele ser bebedor de cerveza.
 - Un bebedor de cerveza suele tener sobrepeso.
 - Por lo general, un infante de la marina está en buena condición física.
- Si se representa la información en una red de herencia se puede concluir que:
- a) Existen dos razonamientos cancelables.
- b) Juan tiene sobrepeso.
- c) No se puede asegurar que Juan sea capellán.
- d) No se puede asegurar que Juan esté en buena condición física.
12. El componente responsable de la gestión de recursos y planificación de tareas en Hadoop es:
- a) MapReduce.
- b) HDFS.
- c) Hadoop Common.
- d) YARN.
13. ¿Qué algoritmos permiten obtener información de una red?
- a) Índices de centralidad.
- b) Detección de comunidades.
- c) Hypertext Induced Topic Selection (HITS).
- d) PageRank.
14. En el contexto del procesamiento de grandes conjuntos de datos, se puede asegurar que:
- a) MapReduce no es adecuado para el procesamiento de datos no estructurados.
- b) MapReduce solo puede manejar pequeñas cantidades de datos y no escala bien a grandes conjuntos de datos.
- c) MapReduce es un enfoque para procesar datos de forma secuencial en un solo servidor para evitar problemas de concurrencia.
- d) MapReduce divide una tarea en múltiples pasos de map y reduce que se ejecutan de forma secuencial en diferentes servidores para mejorar el rendimiento y la escalabilidad.
15. Una buena práctica de SEO para mejorar el posicionamiento de un sitio web en los motores de búsqueda es:
- a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
- b) Copiar contenido directamente de otros sitios web populares para aumentar la cantidad de páginas indexadas.
- c) Llenar el contenido con palabras clave irrelevantes para aumentar la densidad de palabras clave.
- d) Obtener enlaces de otros sitios web relevantes y de calidad que apunten al sitio.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Considerando las prácticas éticas y legales en el Web Scraping, se puede asegurar que:
 - a) Web Scraping sobre datos personales sin consentimiento es generalmente aceptado si los datos se utilizan con fines de investigación.
 - b) Es importante revisar y respetar los términos de servicio del sitio web, así como las leyes aplicables de protección de datos y derechos de autor, antes de realizar Web Scraping.
 - c) Web Scraping requiere considerar las políticas de `robots.txt` del sitio web objetivo.
 - d) La extracción de datos mediante Web Scraping siempre es legal, independientemente de las leyes locales sobre derechos de autor y privacidad de datos definidos en los sitios web.
2. Con respecto a WordNet y su contribución a las ontologías y la representación del conocimiento en los sistemas de RI, puede afirmarse que:
 - a) En WordNet, una palabra está asociada a un conjunto de sinónimos (synsets), siendo estas palabras intercambiables en un contexto.
 - b) WordNet diferencia claramente entre los significados de palabras según su uso en diferentes contextos, lo que permite aplicaciones avanzadas en desambiguación semántica más allá de los sistemas de recomendación y RI.
 - c) La integración de WordNet en sistemas de RI permite la expansión de consultas y la mejora de la precisión de los resultados al entender mejor el significado de los términos de búsqueda a través de su contexto semántico y las relaciones entre palabras.
 - d) Una limitación de WordNet en la representación del conocimiento es su enfoque en el idioma inglés, lo que plantea desafíos en la aplicación global y la interoperabilidad con sistemas de información multilingües.
3. La computación evolutiva:
 - a) Utiliza algoritmos para buscar soluciones óptimas en grandes espacios de búsqueda, lo que la hace adecuada para problemas de optimización en la RI.
 - b) Es útil solo para problemas de clasificación de documentos y no para otras tareas de RI en general.
 - c) No es aplicable en la RI debido a la complejidad de los algoritmos evolutivos.
 - d) Solo puede manejar conjuntos de datos pequeños y no es escalable a grandes volúmenes de datos.
4. La Web 1.0 se caracteriza por:
 - a) Los propietarios de los sitios web proporcionan contenido de forma periódica.
 - b) Uso de comunidades virtuales para popularizar los sitios web de noticias.
 - c) Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
 - d) Los sitios web se centran en brindar información en lugar de facilitar la colaboración o participación de los usuarios.
5. ¿Qué algoritmos permiten obtener información de una red?
 - a) Índices de centralidad.
 - b) Detección de comunidades.
 - c) Hypertext Induced Topic Selection (HITS).
 - d) PageRank.
6. Al “relajar” el concepto de clique en la detección de comunidades se intenta solucionar:
 - a) La uniformidad de los nodos en términos de grado.
 - b) La necesidad de datos externos para analizar la red.
 - c) La imposibilidad de conectar nodos distantes.
 - d) El solapamiento y la complejidad computacional.
7. El algoritmo de PageRank puede describirse como un procedimiento utilizado para:
 - a) Determinar la velocidad de carga de una página web en un navegador.
 - b) Calcular la relevancia de una página web en función de la cantidad y calidad de los enlaces que apuntan hacia ella.
 - c) Clasificar las páginas web en función de su edad y autoridad.

- d) Prevenir el spam y el contenido no deseado en las páginas web.
8. Un investigador necesita recopilar datos de múltiples sitios web para un estudio académico, pero se enfrenta a varios desafíos al realizar el proceso de Web Scraping de manera ética y legal. ¿Cuál de las siguientes opciones describe mejor uno de los desafíos asociados al proceso de Web Scraping?
- La dificultad para encontrar herramientas de Web Scraping gratuitas y fiables.
 - La disponibilidad limitada de datos en línea que se pueden extraer utilizando técnicas de Web Scraping.
 - La necesidad de estar montado sobre un scrawler que cumpla con todas las políticas.
 - La necesidad de comprender la estructura del sitio web y su código HTML para extraer los datos correctamente.
9. Una buena práctica de SEO para mejorar el posicionamiento de un sitio web en los motores de búsqueda es:
- Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - Copiar contenido directamente de otros sitios web populares para aumentar la cantidad de páginas indexadas.
 - Llenar el contenido con palabras clave irrelevantes para aumentar la densidad de palabras clave.
 - Obtener enlaces de otros sitios web relevantes y de calidad que apunten al sitio.
10. No se considera como técnica para detectar comunidades en una red:
- Analizar la mutualidad de los enlaces.
 - Usar el agrupamiento jerárquico.
 - Utilizar el algoritmo de K-Means.
 - Encontrar cliques de vértices de grado par.
11. Se tiene una red de ingredientes donde cada uno representa un nodo y las aristas simbolizan que los ingredientes forman parte de una misma receta. Se busca mejorar la experiencia culinaria mediante la elaboración de combinaciones de ingredientes más interesantes y creativas, para lo cual se debe:
- No tener en cuenta la centralidad de grado de los ingredientes para no darle mayor importancia a los ingredientes más comunes.
 - No considerar la centralidad de grado de los ingredientes, ya que todas las combinaciones de ingredientes son igualmente válidas.
 - Utilizar la centralidad de grado para identificar los ingredientes más populares en la red y crear combinaciones que incluyan una variedad de ingredientes menos comunes.
 - Utilizar la centralidad de grado para identificar los ingredientes menos conectados en la red y tomarlos en cuenta para su inclusión en las combinaciones.
12. En el contexto del procesamiento de grandes conjuntos de datos, se puede asegurar que:
- MapReduce no es adecuado para el procesamiento de datos no estructurados.
 - MapReduce solo puede manejar pequeñas cantidades de datos y no escala bien a grandes conjuntos de datos.
 - MapReduce es un enfoque para procesar datos de forma secuencial en un solo servidor para evitar problemas de concurrencia.
 - MapReduce divide una tarea en múltiples pasos de map y reduce que se ejecutan de forma secuencial en diferentes servidores para mejorar el rendimiento y la escalabilidad.
13. El propósito de la política de revisitado en los Web Crawlers es:
- Determinar la frecuencia con la que los crawlers deben volver a visitar una URL específica para mantener la información actualizada en el índice de búsqueda.
 - Limitar el acceso de los crawlers a ciertos servidores luego de visitar las páginas alojadas en estos.
 - Establecer reglas sobre el tiempo máximo que los crawlers pueden pasar en un sitio web durante cada visita para evitar sobrecargar los servidores.
 - Definir la prioridad de rastreo de las URLs, determinando el orden en que los crawlers visitan y procesan cada página web.
14. Para contribuir positivamente al posicionamiento orgánico de un sitio web en los motores de búsqueda se puede:
- Crear contenido relevante y de alta calidad que satisfaga las necesidades de información de los usuarios.
 - Obtener enlaces entrantes de otros sitios web de alta autoridad y relevancia temática.
 - Incluir una densidad alta de palabras clave para asegurar que el sitio web aparezca en tantas búsquedas como sea posible.
 - Mejorar la velocidad de carga del sitio web y asegurar que sea *responsive* y fácil de usar en dispositivos móviles.

15. Un sistema cuenta con la siguiente información:

- Juan pertenece a la marina.
- Juan es capellán (sacerdote encargado de una tarea específica fuera de la parroquia).
- Los infantes de la marina suelen ser bebedores de cerveza.
- Un capellán no suele ser bebedor de cerveza.
- Un bebedor de cerveza suele tener sobrepeso.

- Por lo general, un infante de la marina está en buena condición física.

Si se representa la información en una red de herencia se puede concluir que:

- a) Existen dos razonamientos cancelables.
- b) Juan tiene sobrepeso.
- c) No se puede asegurar que Juan sea capellán.
- d) No se puede asegurar que Juan esté en buena condición física.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. En un sistema donde el conocimiento está orientado a objetos se puede asegurar que:
 - a) Los marcos y las bandas son las estructuras utilizadas para la representación del modelo.
 - b) Las funciones de agregación dificultan poder establecer relaciones entre los objetos.
 - c) Se enfatiza la atención a la información de la cual se extrajo el conocimiento.
 - d) No es posible definir especializaciones de los objetos de la vida real dentro del sistema.
2. En la RI en el contexto de Big Data, se puede asegurar que:
 - a) El almacenamiento distribuido es una técnica obsoleta.
 - b) El almacenamiento distribuido centraliza todos los datos en un único servidor para facilitar su gestión y mantenimiento.
 - c) El almacenamiento distribuido reparte los datos en múltiples servidores para mejorar la disponibilidad y la redundancia del sistema.
 - d) El almacenamiento distribuido ofrece ventajas significativas en términos de escalabilidad y rendimiento en comparación con el almacenamiento centralizado.
3. Se puede afirmar que:
 - a) Uno de los desafíos en el procesamiento de Big Data es la capacidad de gestionar y analizar datos provenientes de diversas fuentes y en diferentes formatos de manera eficiente.
 - b) MapReduce es un modelo de procesamiento distribuido utilizado para trabajar con grandes volúmenes de datos.
 - c) Los SRI pueden beneficiarse de MapReduce para mejorar la RI relevante.
 - d) El término “Big Data” se refiere exclusivamente al volumen de datos que una organización maneja, sin tener en cuenta la velocidad, la variedad y la veracidad de los datos.
4. En el contexto del procesamiento de grandes conjuntos de datos, se puede asegurar que:
 - a) MapReduce no es adecuado para el procesamiento de datos no estructurados.
 - b) MapReduce solo puede manejar pequeñas cantidades de datos y no escala bien a grandes conjuntos de datos.
 - c) MapReduce es un enfoque para procesar datos de forma secuencial en un solo servidor para evitar problemas de concurrencia.
 - d) MapReduce divide una tarea en múltiples pasos de map y reduce que se ejecutan de forma secuencial en diferentes servidores para mejorar el rendimiento y la escalabilidad.
5. El componente responsable de la gestión de recursos y planificación de tareas en Hadoop es:
 - a) MapReduce.
 - b) HDFS.
 - c) Hadoop Common.
 - d) YARN.
6. En el contexto de la RI en redes, se puede afirmar que:
 - a) La detección de comunidades ayuda a identificar grupos de nodos altamente conectados entre sí, lo que puede ser útil para comprender la estructura y el contenido de la red.
 - b) La detección de comunidades en una red siempre produce resultados objetivos y consistentes independientemente del algoritmo utilizado.
 - c) La detección de comunidades solo se aplica a redes pequeñas y simples, no a redes grandes y complejas.
 - d) La detección de comunidades no es relevante para la RI en redes, ya que se centra únicamente en la estructura de la red sin considerar el contenido.
7. Una empresa de comercio electrónico necesita mejorar su motor de búsqueda para proporcionar resultados más relevantes a sus usuarios. Actualmente, los resultados de la búsqueda no son precisos y los usuarios a menudo encuentran dificultades para encontrar productos específicos. La empresa está considerando implementar una indexación por conceptos para mejorar la relevancia de los resultados de búsqueda. ¿Qué beneficios podría provocar este cambio?

- a) Permite adaptarse fácilmente a cambios en el vocabulario y la terminología utilizada en las descripciones de los productos.
 - b) Facilita la visualización de productos al agruparlos por categorías o características comunes.
 - c) Mejora la precisión en la búsqueda de productos relacionados, incluso cuando no coinciden exactamente con los términos de búsqueda del usuario.
 - d) Ayuda a identificar automáticamente relaciones entre productos, lo que puede mejorar las recomendaciones personalizadas a los usuarios.
8. El concepto de *rank sink* en el algoritmo de PageRank representa páginas web:
 - a) Con una baja calidad de contenido y una cantidad insuficiente de enlaces salientes, lo que las hace menos relevantes en los resultados de búsqueda.
 - b) Con distribuciones uniformes de la puntuación de PageRank.
 - c) Con un alto número de enlaces salientes que no reciben enlaces entrantes, lo que puede afectar negativamente su puntuación de PageRank.
 - d) Que tienen una puntuación de PageRank más alta que otras debido a la manipulación de enlaces entrantes y salientes.
 9. El algoritmo de PageRank converge si:
 - a) Se define un factor de normalización en la fórmula de la función.
 - b) La norma de la diferencia entre los vectores es menor a un umbral predefinido.
 - c) Finaliza la ejecución cuando el número de iteraciones excede un máximo de iteraciones previamente definido.
 - d) El algoritmo no se implementa de forma iterativa.
 10. La política de ordenación de URLs en los Web Crawlers tiene como aspecto fundamental:
 - a) Limitar el acceso de los crawlers a ciertas secciones de un sitio web, evitando el rastreo de URLs consideradas menos importantes o sensibles.
 - b) Establecer la prioridad de rastreo de las URLs, determinando el orden en que los crawlers visitan y procesan cada página web.
 - c) Definir la estructura de la URL de destino, asegurando que estén ordenadas alfabéticamente para facilitar la navegación y la indexación.
 - d) Determinar la forma en que los crawlers asignan un valor de relevancia a cada URL para clasificarlas en el índice de búsqueda.
 11. Se tiene un grafo donde cada nodo es un personaje de cierto libro de cuentos y la existencia de las aristas está definida si dos personajes aparecen en el mismo cuento. Se puede asegurar que:
 - a) La centralidad de grado es útil para identificar los personajes con más conexiones dentro de la red, lo que puede indicar su importancia en el libro.
 - b) La centralidad de cercanía indica el grado de conexión de cada personaje con el resto de los personajes de los cuentos del libro.
 - c) La centralidad de intermediación identifica a los personajes que actúan como conectores entre personajes de cuentos distintos.
 - d) La centralidad de cercanía ofrece una relación entre la cantidad de vecinos de un nodo con respecto a la longitud máxima de un camino dentro del grafo partiendo del nodo en cuestión.
 12. En una red de transporte donde cada nodo es una parada de autobús y las aristas representan si existe un carro que pasa por ambos sitios, ¿qué puede mejorar el sistema de transporte?
 - a) Utilizar la centralidad de intermediación para identificar las paradas de transferencia clave y establecer nuevas conexiones entre diferentes líneas de transporte público.
 - b) Utilizar el grafo inducido de los nodos con mayor valor en la centralidad de grado para aplicar la centralidad de intermediación con el propósito de reforzar las paradas con mayor tráfico.
 - c) No implementar sistemas de información en tiempo real para los usuarios, ya que pueden aumentar la carga de los trabajadores.
 - d) No considerar la centralidad de intermediación de las estaciones de transporte público al planificar rutas y horarios, ya que no tiene impacto en las conexiones entre las paradas.
 13. En el contexto de la representación del conocimiento basada en herencia, ¿qué caracteriza a la herencia cancelable?
 - a) Las propiedades heredadas siempre se mantienen y no pueden anularse.
 - b) La herencia es el resultado del razonamiento no transitivo.
 - c) Las conclusiones no están determinadas y dependen del nodo de interés.
 - d) No existe ambigüedad en las conclusiones obtenidas.
 14. Considerando las prácticas éticas y legales en el Web Scraping, se puede asegurar que:
 - a) Web Scraping sobre datos personales sin consentimiento es generalmente aceptado si los datos se utilizan con fines de investigación.

- b) Es importante revisar y respetar los términos de servicio del sitio web, así como las leyes aplicables de protección de datos y derechos de autor, antes de realizar Web Scraping.
 - c) Web Scraping requiere considerar las políticas de `robots.txt` del sitio web objetivo.
 - d) La extracción de datos mediante Web Scraping siempre es legal, independientemente de las leyes locales sobre derechos de autor y privacidad de datos definidos en los sitios web.
15. En un grafo una comunidad es:
- a) Un conjunto de nodos altamente conectados que forman un subgrafo completamente independiente del resto de la red.
 - b) Un conjunto de nodos aislados.
 - c) Un conjunto de nodos que no comparten ninguna similitud estructural o funcional entre sí.
 - d) Un conjunto de nodos que están más densamente interconectados entre sí que con los nodos fuera del conjunto.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. La diferencia entre la indexación por tokens y la indexación por conceptos puede definirse como:
 - a) La indexación por tokens normaliza los datos reduciéndolos a su forma básica, mientras que la indexación por conceptos utiliza un método de ordenación para organizar los términos característicos de los datos.
 - b) La indexación por tokens asigna un valor numérico a cada término de los datos, mientras que la indexación por conceptos utiliza algoritmos de encriptación para proteger la privacidad de los datos.
 - c) La indexación por tokens asigna pesos a los términos basados en su importancia relativa, mientras que la indexación por conceptos utiliza un sistema de etiquetado para asociar términos con características generales.
 - d) La indexación por tokens divide los datos en términos individuales, mientras que la indexación por conceptos agrupa los datos en categorías definidas.
2. En el modelo de representación del conocimiento basado en herencia se puede asegurar que:
 - a) La cantidad de padres que puede tener un nodo no es mayor que 4.
 - b) Solo se usa en entornos referentes a la biología.
 - c) El razonamiento deducido está respaldado por al menos un camino dentro del grafo.
 - d) Las conclusiones pueden ser canceladas si el grafo es ambiguo.
3. Sobre el algoritmo de PageRank, visto en clase, se puede afirmar que:
 - a) Evalúa la importancia de un sitio web en función de la calidad y cantidad de enlaces entrantes que recibe de otros sitios web.
 - b) Asigna una puntuación baja a los sitios web que contienen muchos enlaces salientes, ya que indica una falta de relevancia.
 - c) Asigna una puntuación alta a los sitios web que tienen un gran número de enlaces entrantes sin tener en cuenta la calidad de esos enlaces.
 - d) Solo tiene en cuenta el contenido en un sitio web para determinar su relevancia en los resultados de búsqueda.
4. Para la RI, el análisis de las redes puede:
 - a) Ayudar a identificar grupos de interés.
 - b) Indicar la importancia de una entidad en la transmisión de la información.
 - c) Revelar patrones de influencia dentro de una comunidad.
 - d) Ayudar a comprender la conectividad y la accesibilidad entre las entidades.
5. Referente al Web Crawler puede afirmarse que:
 - a) Los hipervínculos encontrados en cada sitio web que no pertenecen al dominio donde fueron encontrados se desechan, puesto que no expande el conjunto de URLs sin visitar.
 - b) Las páginas visitadas no se procesan nunca más.
 - c) No necesita de un conjunto inicial de URLs para recorrer la Web.
 - d) No tiene como objetivo indexar y recopilar información de diferentes sitios web.
6. Sobre WordNet y su aplicación en el procesamiento del lenguaje natural, se puede afirmar que:
 - a) Aunque WordNet es una herramienta valiosa en el NLP, su estructura no incluye información sobre la frecuencia de uso de las palabras en el lenguaje natural.
 - b) El diseño de WordNet facilita su integración en aplicaciones multilingües de NLP, aunque su desarrollo original se centró en el inglés.
 - c) Los synsets facilitan la identificación de relaciones semánticas entre palabras, como la hiperonimia y la meronimia, enriqueciendo tareas de NLP.
 - d) WordNet proporciona una base para la desambiguación semántica al agrupar palabras con significados similares en synsets.
7. En una biblioteca digital se necesita implementar un sistema que permita a los usuarios encontrar libros y artículos científicos de forma rápida y precisa. Los documentos están en diversos formatos, incluyendo

- PDF, EPUB y HTML. Se requiere seleccionar un algoritmo de indexación adecuado para el sistema, por lo que el programador designado para la implementación debe considerar:
- La capacidad para manejar documentos en diferentes formatos.
 - La velocidad de indexación y recuperación de datos.
 - La complejidad del algoritmo en términos de implementación y mantenimiento.
 - La capacidad del algoritmo para procesar imágenes incrustadas o referenciadas en los ficheros.
8. Se tiene una red de ingredientes donde cada uno representa un nodo y las aristas simbolizan que los ingredientes forman parte de una misma receta. Se busca mejorar la experiencia culinaria mediante la elaboración de combinaciones de ingredientes más interesantes y creativas, para lo cual se debe:
- No tener en cuenta la centralidad de grado de los ingredientes para no darle mayor importancia a los ingredientes más comunes.
 - No considerar la centralidad de grado de los ingredientes, ya que todas las combinaciones de ingredientes son igualmente válidas.
 - Utilizar la centralidad de grado para identificar los ingredientes más populares en la red y crear combinaciones que incluyan una variedad de ingredientes menos comunes.
 - Utilizar la centralidad de grado para identificar los ingredientes menos conectados en la red y tomarlos en cuenta para su inclusión en las combinaciones.
9. El posicionamiento de un sitio web en los motores de búsqueda puede ser afectado por:
- Mantener una estructura de URL clara y coherente.
 - Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - Contenido sin valor en el sitio web.
 - Obtener enlaces de sitios web irrelevantes y de baja calidad.
10. Un investigador necesita recopilar datos de múltiples sitios web para un estudio académico, pero se enfrenta a varios desafíos al realizar el proceso de Web Scraping de manera ética y legal. ¿Cuál de las siguientes opciones describe mejor uno de los desafíos asociados al proceso de Web Scraping?
- La dificultad para encontrar herramientas de Web Scraping gratuitas y fiables.
 - La disponibilidad limitada de datos en línea que se pueden extraer utilizando técnicas de Web Scraping.
 - La necesidad de estar montado sobre un crawler que cumpla con todas las políticas.
 - La necesidad de comprender la estructura del sitio web y su código HTML para extraer los datos correctamente.
11. El conjunto frontera de URLs en un Web Crawler:
- Delimita los sitios web que visitará en futuras iteraciones del proceso.
 - Indica las secciones que no pueden ser visitadas de cada sitio web.
 - Almacena hipervínculos que pertenecen al mismo dominio del conjunto semilla de URLs con que inició el crawler.
 - Es similar a un conjunto de URLs que esperan ser visitadas.
12. La integración de Hadoop y MapReduce en la RI trae como ventaja:
- La posibilidad de la extracción de información relevante y la generación de resultados significativos de grandes conjuntos de datos.
 - La reducción de los costos operativos a cero.
 - La eliminación de la necesidad de sistemas de bases de datos.
 - La garantía de la privacidad absoluta de los datos procesados.
13. Un sistema cuenta con la siguiente información:
- Juan pertenece a la marina.
 - Juan es capellán (sacerdote encargado de una tarea específica fuera de la parroquia).
 - Los infantes de la marina suelen ser bebedores de cerveza.
 - Un capellán no suele ser bebedor de cerveza.
 - Un bebedor de cerveza suele tener sobrepeso.
 - Por lo general, un infante de la marina está en buena condición física.
- Si se representa la información en una red de herencia se puede concluir que:
- Existen dos razonamientos cancelables.
 - Juan tiene sobrepeso.
 - No se puede asegurar que Juan sea capellán.
 - No se puede asegurar que Juan esté en buena condición física.
14. En la RI en el contexto de Big Data, se puede asegurar que:
- El almacenamiento distribuido es una técnica obsoleta.

- b) El almacenamiento distribuido centraliza todos los datos en un único servidor para facilitar su gestión y mantenimiento.
- c) El almacenamiento distribuido reparte los datos en múltiples servidores para mejorar la disponibilidad y la redundancia del sistema.
- d) El almacenamiento distribuido ofrece ventajas significativas en términos de escalabilidad y rendimiento en comparación con el almacenamiento centralizado.
15. Analizar una red permite:
- a) Evaluar la calidad del contenido de un sitio web.
- b) Obtener predicciones exactas de eventos futuros en mercados financieros.
- c) Detectar posibles tendencias antes de que se conviertan en tendencia.
- d) Encontrar nodos “sensibles” o críticos para la red.

Trabajo de Control
Sistemas de Recuperación de Información
Fecha: 1 de abril de 2024
Temario #8



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Dentro del ecosistema de Hadoop, el HDFS se caracteriza por:
 - a) La capacidad ilimitada de almacenamiento.
 - b) La tolerancia a fallos mediante la replicación de datos.
 - c) El almacenamiento exclusivo para archivos de texto.
 - d) El modelo de acceso y de escritura de datos en tiempo real.
2. Sobre el algoritmo de PageRank, visto en clase, se puede afirmar que:
 - a) Evalúa la importancia de un sitio web en función de la calidad y cantidad de enlaces entrantes que recibe de otros sitios web.
 - b) Asigna una puntuación baja a los sitios web que contienen muchos enlaces salientes, ya que indica una falta de relevancia.
 - c) Asigna una puntuación alta a los sitios web que tienen un gran número de enlaces entrantes sin tener en cuenta la calidad de esos enlaces.
 - d) Solo tiene en cuenta el contenido en un sitio web para determinar su relevancia en los resultados de búsqueda.
3. Una empresa de comercio electrónico necesita mejorar su motor de búsqueda para proporcionar resultados más relevantes a sus usuarios. Actualmente, los resultados de la búsqueda no son precisos y los usuarios a menudo encuentran dificultades para encontrar productos específicos. La empresa está considerando implementar una indexación por conceptos para mejorar la relevancia de los resultados de búsqueda. ¿Qué beneficios podría provocar este cambio?
 - a) Permite adaptarse fácilmente a cambios en el vocabulario y la terminología utilizada en las descripciones de los productos.
 - b) Facilita la visualización de productos al agruparlos por categorías o características comunes.
 - c) Mejora la precisión en la búsqueda de productos relacionados, incluso cuando no coinciden exactamente con los términos de búsqueda del usuario.
 - d) Ayuda a identificar automáticamente relaciones entre productos, lo que puede mejorar las recomendaciones personalizadas a los usuarios.
4. Para contribuir positivamente al posicionamiento orgánico de un sitio web en los motores de búsqueda se puede:
 - a) Crear contenido relevante y de alta calidad que satisfaga las necesidades de información de los usuarios.
 - b) Obtener enlaces entrantes de otros sitios web de alta autoridad y relevancia temática.
 - c) Incluir una densidad alta de palabras clave para asegurar que el sitio web aparezca en tantas búsquedas como sea posible.
 - d) Mejorar la velocidad de carga del sitio web y asegurar que sea *responsive* y fácil de usar en dispositivos móviles.
5. En una red de transporte donde cada nodo es una parada de autobús y las aristas representan si existe un carro que pasa por ambos sitios, ¿qué puede mejorar el sistema de transporte?
 - a) Utilizar la centralidad de intermediación para identificar las paradas de transferencia clave y establecer nuevas conexiones entre diferentes líneas de transporte público.
 - b) Utilizar el grafo inducido de los nodos con mayor valor en la centralidad de grado para aplicar la centralidad de intermediación con el propósito de reforzar las paradas con mayor tráfico.
 - c) No implementar sistemas de información en tiempo real para los usuarios, ya que pueden aumentar la carga de los trabajadores.
 - d) No considerar la centralidad de intermediación de las estaciones de transporte público al planificar rutas y horarios, ya que no tiene impacto en las conexiones entre las paradas.
6. Para la RI, el análisis de las redes puede:
 - a) Ayudar a identificar grupos de interés.
 - b) Indicar la importancia de una entidad en la transmisión de la información.
 - c) Revelar patrones de influencia dentro de una comunidad.
 - d) Ayudar a comprender la conectividad y la accesibilidad entre las entidades.

7. La Web 1.0 se caracteriza por:
 - a) Los propietarios de los sitios web proporcionan contenido de forma periódica.
 - b) Uso de comunidades virtuales para popularizar los sitios web de noticias.
 - c) Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
 - d) Los sitios web se centran en brindar información en lugar de facilitar la colaboración o participación de los usuarios.
8. En el contexto de la representación del conocimiento basada en herencia, ¿qué caracteriza a la herencia cancelable?
 - a) Las propiedades heredadas siempre se mantienen y no pueden anularse.
 - b) La herencia es el resultado del razonamiento no transitivo.
 - c) Las conclusiones no están determinadas y dependen del nodo de interés.
 - d) No existe ambigüedad en las conclusiones obtenidas.
9. En un sistema donde el conocimiento está definido a partir de reglas se puede asegurar que:
 - a) El nivel de especificidad de las reglas está limitado.
 - b) La representación de conocimiento está basada en la lógica proposicional.
 - c) El razonamiento se activa en cadena hacia delante.
 - d) El orden en que se definen las reglas no altera el razonamiento del sistema.
10. El algoritmo de PageRank converge si:
 - a) Se define un factor de normalización en la fórmula de la función.
 - b) La norma de la diferencia entre los vectores es menor a un umbral predefinido.
 - c) Finaliza la ejecución cuando el número de iteraciones excede un máximo de iteraciones previamente definido.
 - d) El algoritmo no se implementa de forma iterativa.
11. Considerando las prácticas éticas y legales en el Web Scraping, se puede asegurar que:
 - a) Web Scraping sobre datos personales sin consentimiento es generalmente aceptado si los datos se utilizan con fines de investigación.
 - b) Es importante revisar y respetar los términos de servicio del sitio web, así como las leyes aplicables de protección de datos y derechos de autor, antes de realizar Web Scraping.
 - c) Web Scraping requiere considerar las políticas de `robots.txt` del sitio web objetivo.
 - d) La extracción de datos mediante Web Scraping siempre es legal, independientemente de las leyes locales sobre derechos de autor y privacidad de datos definidos en los sitios web.
12. Un SRI es capaz de:
 - a) No necesitar de ningún almacenamiento externo para alojar los índices de los datos.
 - b) Reducir el tiempo de indexación de los datos si utiliza servidores distribuidos dentro de la red para que cada uno ejecute la indexación del mismo conjunto de datos.
 - c) Generar índices invertidos de manera óptima sin considerar el contexto.
 - d) Crear los índices asociados a los datos sin tener que analizar cada dato.
13. En la RI en el contexto de Big Data, se puede asegurar que:
 - a) La indexación distribuida es una técnica obsoleta.
 - b) La indexación distribuida almacena todos los datos en un solo servidor para facilitar su acceso y búsqueda.
 - c) La indexación distribuida divide los datos en múltiples fragmentos que se almacenan en varios nodos para permitir búsquedas paralelas y mejorar la escalabilidad.
 - d) La indexación distribuida no ofrece ventajas en términos de rendimiento y escalabilidad en comparación con la indexación centralizada.
14. En un sistema donde el conocimiento está orientado a objetos se puede asegurar que:
 - a) Los marcos y las bandas son las estructuras utilizadas para la representación del modelo.
 - b) Las funciones de agregación dificultan poder establecer relaciones entre los objetos.
 - c) Se enfatiza la atención a la información de la cual se extrajo el conocimiento.
 - d) No es posible definir especializaciones de los objetos de la vida real dentro del sistema.
15. En el diseño de un videojuego de roles (RPG) en el que los personajes tienen atributos como salud, fuerza y velocidad, ¿cuál de las siguientes opciones representa mejor una implementación de la representación del conocimiento orientado a objetos?
 - a) Los personajes se representan como objetos con atributos de salud, fuerza y velocidad.
 - b) Los personajes se representan como clases con atributos de salud, fuerza y velocidad.
 - c) Los personajes se representan como instancias de una clase con atributos de salud, fuerza y velocidad.
 - d) Los personajes se representan como objetos con atributos de salud, fuerza y velocidad.

- a) Cada personaje se representa como un objeto con propiedades como salud, fuerza y velocidad, y métodos para modificar y consultar estos valores.
- b) Cada personaje se representa como una función que calcula sus atributos en función de su nivel y experiencia.
- c) Cada personaje se representa como una matriz de números que almacena sus valores de atributos.
- d) Cada personaje se representa como una lista de cadenas de texto que describen sus características físicas y habilidades.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Con respecto a WordNet y su contribución a las ontologías y la representación del conocimiento en los sistemas de RI, puede afirmarse que:
 - a) En WordNet, una palabra está asociada a un conjunto de sinónimos (synsets), siendo estas palabras intercambiables en un contexto.
 - b) WordNet diferencia claramente entre los significados de palabras según su uso en diferentes contextos, lo que permite aplicaciones avanzadas en desambiguación semántica más allá de los sistemas de recomendación y RI.
 - c) La integración de WordNet en sistemas de RI permite la expansión de consultas y la mejora de la precisión de los resultados al entender mejor el significado de los términos de búsqueda a través de su contexto semántico y las relaciones entre palabras.
 - d) Una limitación de WordNet en la representación del conocimiento es su enfoque en el idioma inglés, lo que plantea desafíos en la aplicación global y la interoperabilidad con sistemas de información multilingües.
2. En el modelo de representación del conocimiento basado en herencia se puede asegurar que:
 - a) La cantidad de padres que puede tener un nodo no es mayor que 4.
 - b) Solo se usa en entornos referentes a la biología.
 - c) El razonamiento deducido está respaldado por al menos un camino dentro del grafo.
 - d) Las conclusiones pueden ser canceladas si el grafo es ambiguo.
3. El concepto de *rank sink* en el algoritmo de PageRank representa páginas web:
 - a) Con una baja calidad de contenido y una cantidad insuficiente de enlaces salientes, lo que las hace menos relevantes en los resultados de búsqueda.
 - b) Con distribuciones uniformes de la puntuación de PageRank.
 - c) Con un alto número de enlaces salientes que no reciben enlaces entrantes, lo que puede afectar negativamente su puntuación de PageRank.
 - d) Que tienen una puntuación de PageRank más alta que otras debido a la manipulación de enlaces entrantes y salientes.
4. En la RI en el contexto de Big Data, se puede asegurar que:
 - a) La indexación distribuida es una técnica obsoleta.
 - b) La indexación distribuida almacena todos los datos en un solo servidor para facilitar su acceso y búsqueda.
 - c) La indexación distribuida divide los datos en múltiples fragmentos que se almacenan en varios nodos para permitir búsquedas paralelas y mejorar la escalabilidad.
 - d) La indexación distribuida no ofrece ventajas en términos de rendimiento y escalabilidad en comparación con la indexación centralizada.
5. Los algoritmos para detectar comunidades en una red intentan:
 - a) Buscar subgrafos tal que no incluyan nodos cuya ausencia desconecte al subgrafo.
 - b) Seleccionar aleatoriamente nodos de alto grado y sus vecinos.
 - c) Buscar conjuntos donde cada nodo de un mismo conjunto tenga características similares al resto de los nodos del conjunto.
 - d) Encontrar grupos donde los nodos pertenecientes a los mismos grupos son cercanos bajo cierta métrica y lejanos con respecto a los nodos de otros grupos.
6. Un sistema cuenta con la siguiente información:
 - Juan pertenece a la marina.
 - Juan es capellán (sacerdote encargado de una tarea específica fuera de la parroquia).
 - Los infantes de la marina suelen ser bebedores de cerveza.
 - Un capellán no suele ser bebedor de cerveza.
 - Un bebedor de cerveza suele tener sobrepeso.
 - Por lo general, un infante de la marina está en buena condición física.

- Si se representa la información en una red de herencia se puede concluir que:
- Existen dos razonamientos cancelables.
 - Juan tiene sobrepeso.
 - No se puede asegurar que Juan sea capellán.
 - No se puede asegurar que Juan esté en buena condición física.
7. Se puede afirmar que:
- Uno de los desafíos en el procesamiento de Big Data es la capacidad de gestionar y analizar datos provenientes de diversas fuentes y en diferentes formatos de manera eficiente.
 - MapReduce es un modelo de procesamiento distribuido utilizado para trabajar con grandes volúmenes de datos.
 - Los SRI pueden beneficiarse de MapReduce para mejorar la RI relevante.
 - El término “Big Data” se refiere exclusivamente al volumen de datos que una organización maneja, sin tener en cuenta la velocidad, la variedad y la veracidad de los datos.
8. En un sistema donde el conocimiento está definido a partir de reglas se puede asegurar que:
- El nivel de especificidad de las reglas está limitado.
 - La representación de conocimiento está basada en la lógica proposicional.
 - El razonamiento se activa en cadena hacia delante.
 - El orden en que se definen las reglas no altera el razonamiento del sistema.
9. Referente al Web Crawler puede afirmarse que:
- Los hipervínculos encontrados en cada sitio web que no pertenecen al dominio donde fueron encontrados se desechan, puesto que no expande el conjunto de URLs sin visitar.
 - Las páginas visitadas no se procesan nunca más.
 - No necesita de un conjunto inicial de URLs para recorrer la Web.
 - No tiene como objetivo indexar y recopilar información de diferentes sitios web.
10. El algoritmo de PageRank puede describirse como un procedimiento utilizado para:
- Determinar la velocidad de carga de una página web en un navegador.
 - Calcular la relevancia de una página web en función de la cantidad y calidad de los enlaces que apuntan hacia ella.
 - Clasificar las páginas web en función de su edad y autoridad.
 - Prevenir el spam y el contenido no deseado en las páginas web.
11. El propósito de la política de revisitado en los Web Crawlers es:
- Determinar la frecuencia con la que los crawlers deben volver a visitar una URL específica para mantener la información actualizada en el índice de búsqueda.
 - Limitar el acceso de los crawlers a ciertos servidores luego de visitar las páginas alojadas en estos.
 - Establecer reglas sobre el tiempo máximo que los crawlers pueden pasar en un sitio web durante cada visita para evitar sobrecargar los servidores.
 - Definir la prioridad de rastreo de las URLs, determinando el orden en que los crawlers visitan y procesan cada página web.
12. ¿Qué estrategia utilizan los Web Crawlers para asegurar un rastreo eficiente y respetuoso de los recursos de los sitios web?
- Extraer únicamente contenido multimedia para reducir la carga en los servidores web.
 - Visitar y rastrear todos los enlaces de una página web simultáneamente.
 - Ignorar completamente el archivo `robots.txt` de los sitios web.
 - Seguir las directrices del archivo `robots.txt` y aplicar un retraso entre las solicitudes.
13. Al realizar la optimización de contenido para SEO debe considerarse:
- Utilizar etiquetas de título y meta descripciones únicas y relevantes para cada página.
 - Seleccionar las palabras clave al azar.
 - Crear contenido valioso y original que satisfaga las necesidades de los usuarios.
 - Incluir palabras clave de manera excesiva en el contenido para mejorar el posicionamiento.
14. El uso de Hadoop y MapReduce en el contexto de la RI tiene como objetivo:
- Facilitar el análisis en tiempo real de datos de redes sociales.
 - Automatizar el mantenimiento de sistemas de bases de datos.
 - Mejorar la eficiencia energética en centros de datos.
 - Procesar y analizar grandes conjuntos de datos para la RI.

15. Se tiene una red de ingredientes donde cada uno representa un nodo y las aristas simbolizan que los ingredientes forman parte de una misma receta. Se busca mejorar la experiencia culinaria mediante la elaboración de combinaciones de ingredientes más interesantes y creativas, para lo cual se debe:
- a) No tener en cuenta la centralidad de grado de los ingredientes para no darle mayor importancia a los ingredientes más comunes.
 - b) No considerar la centralidad de grado de los ingredientes, ya que todas las combinaciones de ingredientes son igualmente válidas.
 - c) Utilizar la centralidad de grado para identificar los ingredientes más populares en la red y crear combinaciones que incluyan una variedad de ingredientes menos comunes.
 - d) Utilizar la centralidad de grado para identificar los ingredientes menos conectados en la red y tomarlos en cuenta para su inclusión en las combinaciones.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. El uso de Hadoop y MapReduce en el contexto de la RI tiene como objetivo:
 - a) Facilitar el análisis en tiempo real de datos de redes sociales.
 - b) Automatizar el mantenimiento de sistemas de bases de datos.
 - c) Mejorar la eficiencia energética en centros de datos.
 - d) Procesar y analizar grandes conjuntos de datos para la RI.
2. Dentro del ecosistema de Hadoop, el HDFS se caracteriza por:
 - a) La capacidad ilimitada de almacenamiento.
 - b) La tolerancia a fallos mediante la replicación de datos.
 - c) El almacenamiento exclusivo para archivos de texto.
 - d) El modelo de acceso y de escritura de datos en tiempo real.
3. Un SRI es capaz de:
 - a) No necesitar de ningún almacenamiento externo para alojar los índices de los datos.
 - b) Reducir el tiempo de indexación de los datos si utiliza servidores distribuidos dentro de la red para que cada uno ejecute la indexación del mismo conjunto de datos.
 - c) Generar índices invertidos de manera óptima sin considerar el contexto.
 - d) Crear los índices asociados a los datos sin tener que analizar cada dato.
4. Al realizar la optimización de contenido para SEO debe considerarse:
 - a) Utilizar etiquetas de título y meta descripciones únicas y relevantes para cada página.
 - b) Seleccionar las palabras clave al azar.
 - c) Crear contenido valioso y original que satisfaga las necesidades de los usuarios.
 - d) Incluir palabras clave de manera excesiva en el contenido para mejorar el posicionamiento.
5. El algoritmo de PageRank converge si:
 - a) Se define un factor de normalización en la fórmula de la función.
 - b) La norma de la diferencia entre los vectores es menor a un umbral predefinido.
 - c) Finaliza la ejecución cuando el número de iteraciones excede un máximo de iteraciones previamente definido.
 - d) El algoritmo no se implementa de forma iterativa.
6. El conjunto frontera de URLs en un Web Crawler:
 - a) Delimita los sitios web que visitará en futuras iteraciones del proceso.
 - b) Indica las secciones que no pueden ser visitadas de cada sitio web.
 - c) Almacena hipervínculos que pertenecen al mismo dominio del conjunto semilla de URLs con que inició el crawler.
 - d) Es similar a un conjunto de URLs que esperan ser visitadas.
7. Se tiene un grafo donde cada nodo es un personaje de cierto libro de cuentos y la existencia de las aristas está definida si dos personajes aparecen en el mismo cuento. Se puede asegurar que:
 - a) La centralidad de grado es útil para identificar los personajes con más conexiones dentro de la red, lo que puede indicar su importancia en el libro.
 - b) La centralidad de cercanía indica el grado de conexión de cada personaje con el resto de los personajes de los cuentos del libro.
 - c) La centralidad de intermediación identifica a los personajes que actúan como conectores entre personajes de cuentos distintos.
 - d) La centralidad de cercanía ofrece una relación entre la cantidad de vecinos de un nodo con respecto a la longitud máxima de un camino dentro del grafo partiendo del nodo en cuestión.
8. Implementar índices invertidos en un SRI asegura:
 - a) Incrementar la seguridad de los datos almacenados al dificultar el acceso directo a la información sin el índice correcto.

- b) Reducir la cantidad de espacio de almacenamiento necesario al comprimir los datos de los documentos.
 - c) Acelerar el proceso de búsqueda al permitir búsquedas directas por contenido en lugar de por título.
 - d) Facilitar la búsqueda sobre los datos que contienen términos específicos al mantener una lista de datos para cada término único.
9. La Web 2.0 se caracteriza por:
- a) Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
 - b) Las plataformas ofrecen experiencias a la medida, permitiendo a los usuarios personalizar sus perfiles, recibir recomendaciones ajustadas al contenido y participar en filtrado colaborativo.
 - c) La existencia de sitios web dinámicos e interactivos que permiten a los usuarios participar, comentar e interactuar tanto con los creadores de contenido como con otros usuarios.
 - d) El uso de distintas tecnologías para crear experiencias web más interactivas y con mayor capacidad de respuesta.
10. Los Web Crawlers se enfrentan a desafíos constantes. Dentro de ellos se encuentran:
- a) La falta de acceso a la base de datos del servidor web para extraer información actualizada.
 - b) La modificación del código y la estructura del sitio web.
 - c) La incapacidad para interpretar correctamente el lenguaje de programación utilizado en el desarrollo de los sitios web.
 - d) La dificultad para generar el contenido dinámico en tiempo real.
11. Se puede afirmar que:
- a) La cantidad de conexiones de un nodo siempre indica su influencia en la red.
 - b) Todas las relaciones en una red tienen la misma importancia para el análisis.
 - c) No es posible identificar subgrupos dentro de una red utilizando análisis de redes.
 - d) El tamaño de una red es siempre indicativo de su efectividad en la transmisión de información.
12. Se puede afirmar que:
- a) Uno de los desafíos en el procesamiento de Big Data es la capacidad de gestionar y analizar datos provenientes de diversas fuentes y en diferentes formatos de manera eficiente.
- b) MapReduce es un modelo de procesamiento distribuido utilizado para trabajar con grandes volúmenes de datos.
 - c) Los SRI pueden beneficiarse de MapReduce para mejorar la RI relevante.
 - d) El término “Big Data” se refiere exclusivamente al volumen de datos que una organización maneja, sin tener en cuenta la velocidad, la variedad y la veracidad de los datos.
13. La afirmación que mejor describe la política de amabilidad en los Web Crawlers es:
- a) Los Web Crawlers son libres de recopilar datos de cualquier sitio web sin restricciones, independientemente de la cantidad de tráfico que generen.
 - b) La política de amabilidad establece pautas y reglas sobre cómo los crawlers deben interactuar con los sitios web para minimizar la carga del servidor y respetar las directivas de los administradores del sitio.
 - c) La política de amabilidad de los Web Crawlers dicta que los crawlers deben priorizar ciertos tipos de contenido sobre otros, ignorando completamente ciertas páginas web.
 - d) Los crawlers se diseñan para acceder a sitios web sin restricciones y extraer datos de manera agresiva para su indexación.
14. La premisa básica del algoritmo de PageRank para clasificar páginas web en los resultados de búsqueda es:
- a) Los enlaces entrantes a una página web desde otras páginas contribuyen a su importancia y clasificación.
 - b) La antigüedad de la página web es el principal factor para determinar su clasificación.
 - c) La cantidad de visitas que recibe una página web determina su clasificación.
 - d) El contenido y la relevancia de las palabras clave en la página web determinan su posición.
15. ¿Qué algoritmos permiten obtener información de una red?
- a) Índices de centralidad.
 - b) Detección de comunidades.
 - c) Hypertext Induced Topic Selection (HITS).
 - d) PageRank.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. La integración de Hadoop y MapReduce en la RI trae como ventaja:
 - a) La posibilidad de la extracción de información relevante y la generación de resultados significativos de grandes conjuntos de datos.
 - b) La reducción de los costos operativos a cero.
 - c) La eliminación de la necesidad de sistemas de bases de datos.
 - d) La garantía de la privacidad absoluta de los datos procesados.
2. Para transformar el contenido no estructurado de las páginas web en datos estructurados, el Web Scraping:
 - a) Convierte de forma automática imágenes a texto.
 - b) Interpreta el código JavaScript en tiempo real.
 - c) Extrae información basada en patrones de HTML/CSS.
 - d) Analiza los protocolos de red.
3. Un sistema cuenta con la siguiente información:
 - Juan pertenece a la marina.
 - Juan es capellán (sacerdote encargado de una tarea específica fuera de la parroquia).
 - Los infantes de la marina suelen ser bebedores de cerveza.
 - Un capellán no suele ser bebedor de cerveza.
 - Un bebedor de cerveza suele tener sobrepeso.
 - Por lo general, un infante de la marina está en buena condición física.

Si se representa la información en una red de herencia se puede concluir que:

 - a) Existen dos razonamientos cancelables.
 - b) Juan tiene sobrepeso.
 - c) No se puede asegurar que Juan sea capellán.
 - d) No se puede asegurar que Juan esté en buena condición física.
4. La transición de la Web 1.0 a la Web 2.0 se caracterizó principalmente por:
 - a) La reducción en el uso de HTML y CSS en el desarrollo de sitios web.
 - b) La disminución de la importancia de los motores de búsqueda en la navegación web.
 - c) El cambio de páginas web estáticas a dinámicas, permitiendo la interacción del usuario y la generación de contenido.
 - d) El aumento en la velocidad de conexión a internet, que permitió una mejor calidad de las páginas web.
5. Para contribuir positivamente al posicionamiento orgánico de un sitio web en los motores de búsqueda se puede:
 - a) Crear contenido relevante y de alta calidad que satisfaga las necesidades de información de los usuarios.
 - b) Obtener enlaces entrantes de otros sitios web de alta autoridad y relevancia temática.
 - c) Incluir una densidad alta de palabras clave para asegurar que el sitio web aparezca en tantas búsquedas como sea posible.
 - d) Mejorar la velocidad de carga del sitio web y asegurar que sea *responsive* y fácil de usar en dispositivos móviles.
6. Dentro del ecosistema de Hadoop, el HDFS se caracteriza por:
 - a) La capacidad ilimitada de almacenamiento.
 - b) La tolerancia a fallos mediante la replicación de datos.
 - c) El almacenamiento exclusivo para archivos de texto.
 - d) El modelo de acceso y de escritura de datos en tiempo real.
7. Se puede afirmar que:
 - a) La cantidad de conexiones de un nodo siempre indica su influencia en la red.
 - b) Todas las relaciones en una red tienen la misma importancia para el análisis.
 - c) No es posible identificar subgrupos dentro de una red utilizando análisis de redes.
 - d) El tamaño de una red es siempre indicativo de su efectividad en la transmisión de información.

8. Un investigador necesita recopilar datos de múltiples sitios web para un estudio académico, pero se enfrenta a varios desafíos al realizar el proceso de Web Scraping de manera ética y legal. ¿Cuál de las siguientes opciones describe mejor uno de los desafíos asociados al proceso de Web Scraping?
 - a) La dificultad para encontrar herramientas de Web Scraping gratuitas y fiables.
 - b) La disponibilidad limitada de datos en línea que se pueden extraer utilizando técnicas de Web Scraping.
 - c) La necesidad de estar montado sobre un scrawler que cumpla con todas las políticas.
 - d) La necesidad de comprender la estructura del sitio web y su código HTML para extraer los datos correctamente.
9. Una plataforma de comercio electrónico desea mejorar la experiencia del usuario al permitir una navegación más personalizada y contextualizada. Actualmente los usuarios tienen dificultades para encontrar productos relevantes debido a la gran cantidad de opciones disponibles. La empresa está interesada en implementar características de la Web 2.5 y la Web Semántica para abordar este problema. ¿Qué características podrían ayudar para ofrecer una navegación más personalizada y contextualizada?
 - a) La implementación de ontologías y metadatos para enriquecer la descripción de productos y mejorar la precisión de las recomendaciones.
 - b) La optimización de la velocidad de carga del sitio web para mejorar la experiencia del usuario y reducir el abandono del carrito de compra.
 - c) La incorporación de los productos en tendencia en el mercado.
 - d) La integración de redes sociales para permitir la recomendación de productos basada en las preferencias de amigos y contactos.
10. En el contexto de la representación del conocimiento basada en herencia, ¿qué caracteriza a la herencia cancelable?
 - a) Las propiedades heredadas siempre se mantienen y no pueden anularse.
 - b) La herencia es el resultado del razonamiento no transitivo.
 - c) Las conclusiones no están determinadas y dependen del nodo de interés.
 - d) No existe ambigüedad en las conclusiones obtenidas.
11. El algoritmo de PageRank converge si:
 - a) Se define un factor de normalización en la fórmula de la función.
 - b) La norma de la diferencia entre los vectores es menor a un umbral predefinido.
 - c) Finaliza la ejecución cuando el número de iteraciones excede un máximo de iteraciones previamente definido.
 - d) El algoritmo no se implementa de forma iterativa.
12. En la RI en el contexto de Big Data, se puede asegurar que:
 - a) El almacenamiento distribuido es una técnica obsoleta.
 - b) El almacenamiento distribuido centraliza todos los datos en un único servidor para facilitar su gestión y mantenimiento.
 - c) El almacenamiento distribuido reparte los datos en múltiples servidores para mejorar la disponibilidad y la redundancia del sistema.
 - d) El almacenamiento distribuido ofrece ventajas significativas en términos de escalabilidad y rendimiento en comparación con el almacenamiento centralizado.
13. ¿Cuál de las siguientes opciones describe mejor la diferencia clave entre Web Crawling y Web Scraping?
 - a) Web Scraping implica el análisis de la estructura y el contenido de las páginas web para extraer datos, mientras que Web Crawling se refiere a la descarga y almacenamiento de páginas web completas.
 - b) Web Scraping es más eficaz para rastrear e indexar contenido web para motores de búsqueda, mientras que Web Crawling se utiliza principalmente para la extracción de datos en proyectos de investigación.
 - c) Web Crawling se centra en la exploración y recopilación de enlaces de múltiples sitios web, mientras que Web Scraping se enfoca en la extracción específica de datos de páginas web individuales.
 - d) Web Crawling se realiza utilizando herramientas de automatización como Selenium WebDriver, mientras que Web Scraping se lleva a cabo mediante el análisis de HTML y CSS.
14. Los algoritmos para detectar comunidades en una red intentan:
 - a) Buscar subgrafos tal que no incluyan nodos cuya ausencia desconecte al subgrafo.
 - b) Seleccionar aleatoriamente nodos de alto grado y sus vecinos.
 - c) Buscar conjuntos donde cada nodo de un mismo conjunto tenga características similares al resto de los nodos del conjunto.

- d) Encontrar grupos donde los nodos pertenecientes a los mismos grupos son cercanos bajo cierta métrica y lejanos con respecto a los nodos de otros grupos.
15. Considerando las prácticas éticas y legales en el Web Scraping, se puede asegurar que:
- a) Web Scraping sobre datos personales sin consentimiento es generalmente aceptado si los datos se utilizan con fines de investigación.
 - b) Es importante revisar y respetar los términos de servicio del sitio web, así como las leyes aplicables de protección de datos y derechos de autor, antes de realizar Web Scraping.
 - c) Web Scraping requiere considerar las políticas de `robots.txt` del sitio web objetivo.
 - d) La extracción de datos mediante Web Scraping siempre es legal, independientemente de las leyes locales sobre derechos de autor y privacidad de datos definidos en los sitios web.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. En un grafo una comunidad es:
 - a) Un conjunto de nodos altamente conectados que forman un subgrafo completamente independiente del resto de la red.
 - b) Un conjunto de nodos aislados.
 - c) Un conjunto de nodos que no comparten ninguna similitud estructural o funcional entre sí.
 - d) Un conjunto de nodos que están más densamente interconectados entre sí que con los nodos fuera del conjunto.
2. La Web 1.0 se caracteriza por:
 - a) Los propietarios de los sitios web proporcionan contenido de forma periódica.
 - b) Uso de comunidades virtuales para popularizar los sitios web de noticias.
 - c) Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
 - d) Los sitios web se centran en brindar información en lugar de facilitar la colaboración o participación de los usuarios.
3. Para transformar el contenido no estructurado de las páginas web en datos estructurados, el Web Scraping:
 - a) Convierte de forma automática imágenes a texto.
 - b) Interpreta el código JavaScript en tiempo real.
 - c) Extrae información basada en patrones de HTML/CSS.
 - d) Analiza los protocolos de red.
4. El algoritmo Hypertext Induced Topic Selection (HITS) intenta buscar nodos especiales. Estos son conocidos como:
 - a) Sensibles.
 - b) Centrales.
 - c) Hubs.
 - d) Autoridades.
5. La computación evolutiva:
 - a) Utiliza algoritmos para buscar soluciones óptimas en grandes espacios de búsqueda, lo que la hace adecuada para problemas de optimización en la RI.
 - b) Es útil solo para problemas de clasificación de documentos y no para otras tareas de RI en general.
 - c) No es aplicable en la RI debido a la complejidad de los algoritmos evolutivos.
 - d) Solo puede manejar conjuntos de datos pequeños y no es escalable a grandes volúmenes de datos.
6. El conjunto frontera de URLs en un Web Crawler:
 - a) Delimita los sitios web que visitará en futuras iteraciones del proceso.
 - b) Indica las secciones que no pueden ser visitadas de cada sitio web.
 - c) Almacena hipervínculos que pertenecen al mismo dominio del conjunto semilla de URLs con que inició el crawler.
 - d) Es similar a un conjunto de URLs que esperan ser visitadas.
7. En una biblioteca digital se necesita implementar un sistema que permita a los usuarios encontrar libros y artículos científicos de forma rápida y precisa. Los documentos están en diversos formatos, incluyendo PDF, EPUB y HTML. Se requiere seleccionar un algoritmo de indexación adecuado para el sistema, por lo que el programador designado para la implementación debe considerar:
 - a) La capacidad para manejar documentos en diferentes formatos.
 - b) La velocidad de indexación y recuperación de datos.
 - c) La complejidad del algoritmo en términos de implementación y mantenimiento.
 - d) La capacidad del algoritmo para procesar imágenes incrustadas o referenciadas en los ficheros.
8. No se considera como técnica para detectar comunidades en una red:
 - a) Analizar la mutualidad de los enlaces.
 - b) Usar el agrupamiento jerárquico.

- c) Utilizar el algoritmo de K-Means.
 - d) Encontrar cliques de vértices de grado par.
9. La política de ordenación de URLs en los Web Crawlers tiene como aspecto fundamental:
- a) Limitar el acceso de los crawlers a ciertas secciones de un sitio web, evitando el rastreo de URLs consideradas menos importantes o sensibles.
 - b) Establecer la prioridad de rastreo de las URLs, determinando el orden en que los crawlers visitan y procesan cada página web.
 - c) Definir la estructura de la URL de destino, asegurando que estén ordenadas alfabéticamente para facilitar la navegación y la indexación.
 - d) Determinar la forma en que los crawlers asignan un valor de relevancia a cada URL para clasificarlas en el índice de búsqueda.
10. El algoritmo de PageRank puede describirse como un procedimiento utilizado para:
- a) Determinar la velocidad de carga de una página web en un navegador.
 - b) Calcular la relevancia de una página web en función de la cantidad y calidad de los enlaces que apuntan hacia ella.
 - c) Clasificar las páginas web en función de su edad y autoridad.
 - d) Prevenir el spam y el contenido no deseado en las páginas web.
11. Un SRI es capaz de:
- a) No necesitar de ningún almacenamiento externo para alojar los índices de los datos.
 - b) Reducir el tiempo de indexación de los datos si utiliza servidores distribuidos dentro de la red para que cada uno ejecute la indexación del mismo conjunto de datos.
 - c) Generar índices invertidos de manera óptima sin considerar el contexto.
 - d) Crear los índices asociados a los datos sin tener que analizar cada dato.
12. En el contexto de la representación del conocimiento basada en herencia, ¿qué caracteriza a la herencia cancelable?
- a) Las propiedades heredadas siempre se mantienen y no pueden anularse.
 - b) La herencia es el resultado del razonamiento no transitivo.
 - c) Las conclusiones no están determinadas y dependen del nodo de interés.
 - d) No existe ambigüedad en las conclusiones obtenidas.
13. Existen varias estrategias consideradas como no recomendables para mejorar el SEO de un sitio web. Dentro de estas estrategias negativas, se encuentran:
- a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Evitar el uso de las meta etiquetas y descripciones del sitio web para reflejar el contenido de la página.
 - c) Construir enlaces naturales de sitios web con autoridad y relevancia temática.
 - d) Incluir una cantidad excesiva de palabras clave irrelevantes para intentar manipular los rankings de búsqueda.
14. ¿Cuál de las siguientes opciones describe mejor la diferencia clave entre Web Crawling y Web Scraping?
- a) Web Scraping implica el análisis de la estructura y el contenido de las páginas web para extraer datos, mientras que Web Crawling se refiere a la descarga y almacenamiento de páginas web completas.
 - b) Web Scraping es más eficaz para rastrear e indexar contenido web para motores de búsqueda, mientras que Web Crawling se utiliza principalmente para la extracción de datos en proyectos de investigación.
 - c) Web Crawling se centra en la exploración y recopilación de enlaces de múltiples sitios web, mientras que Web Scraping se enfoca en la extracción específica de datos de páginas web individuales.
 - d) Web Crawling se realiza utilizando herramientas de automatización como Selenium WebDriver, mientras que Web Scraping se lleva a cabo mediante el análisis de HTML y CSS.
15. En la RI en el contexto de Big Data, se puede asegurar que:
- a) El almacenamiento distribuido es una técnica obsoleta.
 - b) El almacenamiento distribuido centraliza todos los datos en un único servidor para facilitar su gestión y mantenimiento.
 - c) El almacenamiento distribuido reparte los datos en múltiples servidores para mejorar la disponibilidad y la redundancia del sistema.
 - d) El almacenamiento distribuido ofrece ventajas significativas en términos de escalabilidad y rendimiento en comparación con el almacenamiento centralizado.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. El algoritmo Hypertext Induced Topic Selection (HITS) intenta buscar nodos especiales. Estos son conocidos como:
 - a) Sensibles.
 - b) Centrales.
 - c) Hubs.
 - d) Autoridades.
2. La Web 1.0 se caracteriza por:
 - a) Los propietarios de los sitios web proporcionan contenido de forma periódica.
 - b) Uso de comunidades virtuales para popularizar los sitios web de noticias.
 - c) Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
 - d) Los sitios web se centran en brindar información en lugar de facilitar la colaboración o participación de los usuarios.
3. En una empresa comercial, se tiene una red donde los nodos corresponden a los empleados y las aristas representan la frecuencia con la que colaboran en las ventas. Se busca mejorar la colaboración entre los empleados para aumentar las ventas totales, por lo que la directiva debe:
 - a) No tener en cuenta la centralidad de cercanía de los empleados al diseñar estrategias de colaboración, ya que esto podría introducir complicaciones adicionales en el proceso de trabajo.
 - b) Utilizar la centralidad de cercanía para identificar los empleados menos cercanos a otros en la red y asignarles tareas individuales para evitar posibles conflictos y desacuerdos en el proceso de colaboración.
 - c) No considerar la centralidad de cercanía de los empleados al diseñar estrategias de colaboración, ya que todos los empleados tienen el mismo acceso a la información y los recursos.
 - d) Utilizar la centralidad de cercanía para identificar los empleados más cercanos a otros en la red y promover la colaboración entre ellos, facilitando así la comunicación y el intercambio de conocimientos para mejorar las ventas.
4. Dentro del ecosistema de Hadoop, el HDFS se caracteriza por:
 - a) La capacidad ilimitada de almacenamiento.
 - b) La tolerancia a fallos mediante la replicación de datos.
 - c) El almacenamiento exclusivo para archivos de texto.
 - d) El modelo de acceso y de escritura de datos en tiempo real.
5. En el contexto de la representación del conocimiento basada en herencia, ¿qué caracteriza a la herencia cancelable?
 - a) Las propiedades heredadas siempre se mantienen y no pueden anularse.
 - b) La herencia es el resultado del razonamiento no transitivo.
 - c) Las conclusiones no están determinadas y dependen del nodo de interés.
 - d) No existe ambigüedad en las conclusiones obtenidas.
6. El algoritmo de PageRank puede describirse como un procedimiento utilizado para:
 - a) Determinar la velocidad de carga de una página web en un navegador.
 - b) Calcular la relevancia de una página web en función de la cantidad y calidad de los enlaces que apuntan hacia ella.
 - c) Clasificar las páginas web en función de su edad y autoridad.
 - d) Prevenir el spam y el contenido no deseado en las páginas web.
7. Con respecto a WordNet y su contribución a las ontologías y la representación del conocimiento en los sistemas de RI, puede afirmarse que:
 - a) En WordNet, una palabra está asociada a un conjunto de sinónimos (synsets), siendo estas palabras intercambiables en un contexto.
 - b) WordNet diferencia claramente entre los significados de palabras según su uso en diferentes contextos, lo que permite aplicaciones avanzadas en desambiguación semántica más allá de los sistemas de recomendación y RI.

- c) La integración de WordNet en sistemas de RI permite la expansión de consultas y la mejora de la precisión de los resultados al entender mejor el significado de los términos de búsqueda a través de su contexto semántico y las relaciones entre palabras.
 - d) Una limitación de WordNet en la representación del conocimiento es su enfoque en el idioma inglés, lo que plantea desafíos en la aplicación global y la interoperabilidad con sistemas de información multilingües.
8. Se tiene un grafo donde cada nodo es un personaje de cierto libro de cuentos y la existencia de las aristas está definida si dos personajes aparecen en el mismo cuento. Se puede asegurar que:
- a) La centralidad de grado es útil para identificar los personajes con más conexiones dentro de la red, lo que puede indicar su importancia en el libro.
 - b) La centralidad de cercanía indica el grado de conexión de cada personaje con el resto de los personajes de los cuentos del libro.
 - c) La centralidad de intermediación identifica a los personajes que actúan como conectores entre personajes de cuentos distintos.
 - d) La centralidad de cercanía ofrece una relación entre la cantidad de vecinos de un nodo con respecto a la longitud máxima de un camino dentro del grafo partiendo del nodo en cuestión.
9. La integración de Hadoop y MapReduce en la RI trae como ventaja:
- a) La posibilidad de la extracción de información relevante y la generación de resultados significativos de grandes conjuntos de datos.
 - b) La reducción de los costos operativos a cero.
 - c) La eliminación de la necesidad de sistemas de bases de datos.
 - d) La garantía de la privacidad absoluta de los datos procesados.
10. ¿Qué estrategia utilizan los Web Crawlers para asegurar un rastreo eficiente y respetuoso de los recursos de los sitios web?
- a) Extraer únicamente contenido multimedia para reducir la carga en los servidores web.
 - b) Visitar y rastrear todos los enlaces de una página web simultáneamente.
 - c) Ignorar completamente el archivo `robots.txt` de los sitios web.
 - d) Seguir las directrices del archivo `robots.txt` y aplicar un retraso entre las solicitudes.
11. El concepto de *rank sink* en el algoritmo de PageRank representa páginas web:
- a) Con una baja calidad de contenido y una cantidad insuficiente de enlaces salientes, lo que las hace menos relevantes en los resultados de búsqueda.
 - b) Con distribuciones uniformes de la puntuación de PageRank.
 - c) Con un alto número de enlaces salientes que no reciben enlaces entrantes, lo que puede afectar negativamente su puntuación de PageRank.
 - d) Que tienen una puntuación de PageRank más alta que otras debido a la manipulación de enlaces entrantes y salientes.
12. Al realizar la optimización de contenido para SEO debe considerarse:
- a) Utilizar etiquetas de título y meta descripciones únicas y relevantes para cada página.
 - b) Seleccionar las palabras clave al azar.
 - c) Crear contenido valioso y original que satisfaga las necesidades de los usuarios.
 - d) Incluir palabras clave de manera excesiva en el contenido para mejorar el posicionamiento.
13. En un sistema donde el conocimiento está definido a partir de reglas se puede asegurar que:
- a) El nivel de especificidad de las reglas está limitado.
 - b) La representación de conocimiento está basada en la lógica proposicional.
 - c) El razonamiento se activa en cadena hacia delante.
 - d) El orden en que se definen las reglas no altera el razonamiento del sistema.
14. La afirmación que mejor refleja el principio subyacente de PageRank, considerando su importancia en la clasificación de los sitios web, es:
- a) El algoritmo de PageRank considera tanto la cantidad como la calidad de los enlaces entrantes, asignando mayor valor a los enlaces provenientes de sitios web considerados como "importantes".
 - b) PageRank valora más la cantidad de enlaces entrantes a una página web, independientemente de la calidad o relevancia de estos enlaces.
 - c) La efectividad de PageRank se basa exclusivamente en el análisis de las palabras clave contenidas en los enlaces entrantes, sin considerar la estructura de enlace de la Web.
 - d) PageRank opera bajo el supuesto de que los enlaces entrantes y salientes tienen el mismo impacto en la valoración de la relevancia de una página web.

15. En la RI en el contexto de Big Data, se puede asegurar que:
- a) La indexación distribuida es una técnica obsoleta.
 - b) La indexación distribuida almacena todos los datos en un solo servidor para facilitar su acceso y búsqueda.
 - c) La indexación distribuida divide los datos en múltiples fragmentos que se almacenan en varios nodos para permitir búsquedas paralelas y mejorar la escalabilidad.
 - d) La indexación distribuida no ofrece ventajas en términos de rendimiento y escalabilidad en comparación con la indexación centralizada.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. ¿Qué algoritmos permiten obtener información de una red?
 - a) Índices de centralidad.
 - b) Detección de comunidades.
 - c) Hypertext Induced Topic Selection (HITS).
 - d) PageRank.
2. Se tiene un grafo donde cada nodo es un personaje de cierto libro de cuentos y la existencia de las aristas está definida si dos personajes aparecen en el mismo cuento. Se puede asegurar que:
 - a) La centralidad de grado es útil para identificar los personajes con más conexiones dentro de la red, lo que puede indicar su importancia en el libro.
 - b) La centralidad de cercanía indica el grado de conexión de cada personaje con el resto de los personajes de los cuentos del libro.
 - c) La centralidad de intermediación identifica a los personajes que actúan como conectores entre personajes de cuentos distintos.
 - d) La centralidad de cercanía ofrece una relación entre la cantidad de vecinos de un nodo con respecto a la longitud máxima de un camino dentro del grafo partiendo del nodo en cuestión.
3. El uso de Hadoop y MapReduce en el contexto de la RI tiene como objetivo:
 - a) Facilitar el análisis en tiempo real de datos de redes sociales.
 - b) Automatizar el mantenimiento de sistemas de bases de datos.
 - c) Mejorar la eficiencia energética en centros de datos.
 - d) Procesar y analizar grandes conjuntos de datos para la RI.
4. El algoritmo de PageRank converge si:
 - a) Se define un factor de normalización en la fórmula de la función.
 - b) La norma de la diferencia entre los vectores es menor a un umbral predefinido.
 - c) Finaliza la ejecución cuando el número de iteraciones excede un máximo de iteraciones previamente definido.
 - d) El algoritmo no se implementa de forma iterativa.
5. Considerando las dimensiones y desafíos inherentes a Big Data puede afirmarse, tomando en cuenta las características clave y las implicaciones para su procesamiento y análisis, que:
 - a) Big Data se caracteriza principalmente por su pequeño volumen y uniformidad, permitiendo un procesamiento eficiente con mínimas adaptaciones de las herramientas de análisis de datos tradicionales.
 - b) Big Data no desafía la capacidad de las herramientas tradicionales de procesamiento de datos para capturar, almacenar, gestionar y analizar efectivamente la información, dada la evolución constante de las capacidades computacionales y algoritmos de optimización.
 - c) Aunque Big Data puede incluir datos estructurados, su naturaleza se expande al incorporar grandes cantidades de datos no estructurados y semiestructurados, lo que exige el uso de tecnologías especializadas en almacenamiento y procesamiento como Hadoop y sistemas de bases de datos NoSQL.
 - d) Además de su complejidad y diversidad, Big Data introduce desafíos significativos en términos de veracidad y variabilidad al requerir métodos avanzados de limpieza y validación de datos para asegurar la integridad del análisis.
6. La computación evolutiva:
 - a) Utiliza algoritmos para buscar soluciones óptimas en grandes espacios de búsqueda, lo que la hace adecuada para problemas de optimización en la RI.
 - b) Es útil solo para problemas de clasificación de documentos y no para otras tareas de RI en general.
 - c) No es aplicable en la RI debido a la complejidad de los algoritmos evolutivos.
 - d) Solo puede manejar conjuntos de datos pequeños y no es escalable a grandes volúmenes de datos.
7. La afirmación que mejor refleja el principio subyacente de PageRank, considerando su importancia en la clasificación de los sitios web, es:

- a) El algoritmo de PageRank considera tanto la cantidad como la calidad de los enlaces entrantes, asignando mayor valor a los enlaces provenientes de sitios web considerados como “importantes”.
 - b) PageRank valora más la cantidad de enlaces entrantes a una página web, independientemente de la calidad o relevancia de estos enlaces.
 - c) La efectividad de PageRank se basa exclusivamente en el análisis de las palabras clave contenidas en los enlaces entrantes, sin considerar la estructura de enlace de la Web.
 - d) PageRank opera bajo el supuesto de que los enlaces entrantes y salientes tienen el mismo impacto en la valoración de la relevancia de una página web.
8. Considerando las prácticas éticas y legales en el Web Scraping, se puede asegurar que:
- a) Web Scraping sobre datos personales sin consentimiento es generalmente aceptado si los datos se utilizan con fines de investigación.
 - b) Es importante revisar y respetar los términos de servicio del sitio web, así como las leyes aplicables de protección de datos y derechos de autor, antes de realizar Web Scraping.
 - c) Web Scraping requiere considerar las políticas de `robots.txt` del sitio web objetivo.
 - d) La extracción de datos mediante Web Scraping siempre es legal, independientemente de las leyes locales sobre derechos de autor y privacidad de datos definidos en los sitios web.
9. Un SRI es capaz de:
- a) No necesitar de ningún almacenamiento externo para alojar los índices de los datos.
 - b) Reducir el tiempo de indexación de los datos si utiliza servidores distribuidos dentro de la red para que cada uno ejecute la indexación del mismo conjunto de datos.
 - c) Generar índices invertidos de manera óptima sin considerar el contexto.
 - d) Crear los índices asociados a los datos sin tener que analizar cada dato.
10. En un sistema donde el conocimiento está definido a partir de reglas se puede asegurar que:
- a) El nivel de especificidad de las reglas está limitado.
 - b) La representación de conocimiento está basada en la lógica proposicional.
 - c) El razonamiento se activa en cadena hacia delante.
 - d) El orden en que se definen las reglas no altera el razonamiento del sistema.
11. En un sistema de control de tráfico urbano basado en reglas, ¿cuál de las siguientes reglas sería más efectiva para manejar situaciones de congestión en una intersección?
- a) Si hay pocos vehículos en la intersección, reducir el tiempo de los semáforos en verde.
 - b) Si hay muchos vehículos en la intersección, aumentar el tiempo de los semáforos en verde.
 - c) Si hay un vehículo de emergencia en la intersección, detener todos los demás vehículos.
 - d) Si hay muchos peatones cruzando la intersección, reducir el tiempo de los semáforos en rojo.
12. La Web 1.0 se caracteriza por:
- a) Los propietarios de los sitios web proporcionan contenido de forma periódica.
 - b) Uso de comunidades virtuales para popularizar los sitios web de noticias.
 - c) Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
 - d) Los sitios web se centran en brindar información en lugar de facilitar la colaboración o participación de los usuarios.
13. Al realizar la optimización de contenido para SEO debe considerarse:
- a) Utilizar etiquetas de título y meta descripciones únicas y relevantes para cada página.
 - b) Seleccionar las palabras clave al azar.
 - c) Crear contenido valioso y original que satisfaga las necesidades de los usuarios.
 - d) Incluir palabras clave de manera excesiva en el contenido para mejorar el posicionamiento.
14. Si se tiene el conjunto de páginas interconectadas $\{A \rightarrow B, C; B \rightarrow C; C \rightarrow A\}$, entonces la página con valor más alto de PageRank es:
- a) B.
 - b) C.
 - c) A y B.
 - d) A.
15. En la RI en el contexto de Big Data, se puede asegurar que:
- a) La indexación distribuida es una técnica obsoleta.
 - b) La indexación distribuida almacena todos los datos en un solo servidor para facilitar su acceso y búsqueda.

- c) La indexación distribuida divide los datos en múltiples fragmentos que se almacenan en varios nodos para permitir búsquedas paralelas y mejorar la escalabilidad.
- d) La indexación distribuida no ofrece ventajas en términos de rendimiento y escalabilidad en comparación con la indexación centralizada.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. En un grafo una comunidad es:
 - a) Un conjunto de nodos altamente conectados que forman un subgrafo completamente independiente del resto de la red.
 - b) Un conjunto de nodos aislados.
 - c) Un conjunto de nodos que no comparten ninguna similitud estructural o funcional entre sí.
 - d) Un conjunto de nodos que están más densamente interconectados entre sí que con los nodos fuera del conjunto.
2. El concepto de *rank sink* en el algoritmo de PageRank representa páginas web:
 - a) Con una baja calidad de contenido y una cantidad insuficiente de enlaces salientes, lo que las hace menos relevantes en los resultados de búsqueda.
 - b) Con distribuciones uniformes de la puntuación de PageRank.
 - c) Con un alto número de enlaces salientes que no reciben enlaces entrantes, lo que puede afectar negativamente su puntuación de PageRank.
 - d) Que tienen una puntuación de PageRank más alta que otras debido a la manipulación de enlaces entrantes y salientes.
3. ¿Qué estrategia utilizan los Web Crawlers para asegurar un rastreo eficiente y respetuoso de los recursos de los sitios web?
 - a) Extraer únicamente contenido multimedia para reducir la carga en los servidores web.
 - b) Visitar y rastrear todos los enlaces de una página web simultáneamente.
 - c) Ignorar completamente el archivo `robots.txt` de los sitios web.
 - d) Seguir las directrices del archivo `robots.txt` y aplicar un retraso entre las solicitudes.
4. Sobre WordNet y su aplicación en el procesamiento del lenguaje natural, se puede afirmar que:
 - a) Aunque WordNet es una herramienta valiosa en el NLP, su estructura no incluye información sobre la frecuencia de uso de las palabras en el lenguaje natural.
 - b) El diseño de WordNet facilita su integración en aplicaciones multilingües de NLP, aunque su desarrollo original se centró en el inglés.
 - c) Los synsets facilitan la identificación de relaciones semánticas entre palabras, como la hiperonimia y la meronimia, enriqueciendo tareas de NLP.
 - d) WordNet proporciona una base para la desambiguación semántica al agrupar palabras con significados similares en synsets.
5. La afirmación que mejor describe la política de amabilidad en los Web Crawlers es:
 - a) Los Web Crawlers son libres de recopilar datos de cualquier sitio web sin restricciones, independientemente de la cantidad de tráfico que generen.
 - b) La política de amabilidad establece pautas y reglas sobre cómo los crawlers deben interactuar con los sitios web para minimizar la carga del servidor y respetar las directivas de los administradores del sitio.
 - c) La política de amabilidad de los Web Crawlers dicta que los crawlers deben priorizar ciertos tipos de contenido sobre otros, ignorando completamente ciertas páginas web.
 - d) Los crawlers se diseñan para acceder a sitios web sin restricciones y extraer datos de manera agresiva para su indexación.
6. El algoritmo de PageRank puede describirse como un procedimiento utilizado para:
 - a) Determinar la velocidad de carga de una página web en un navegador.
 - b) Calcular la relevancia de una página web en función de la cantidad y calidad de los enlaces que apuntan hacia ella.
 - c) Clasificar las páginas web en función de su edad y autoridad.
 - d) Prevenir el spam y el contenido no deseado en las páginas web.
7. El posicionamiento de un sitio web en los motores de búsqueda puede ser afectado por:
 - a) Mantener una estructura de URL clara y coherente.

- b) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - c) Contenido sin valor en el sitio web.
 - d) Obtener enlaces de sitios web irrelevantes y de baja calidad.
8. En el contexto de la RI en redes sociales se puede afirmar que:
- a) El análisis de redes sociales se centra exclusivamente en la cantidad de seguidores que tiene un usuario en particular para determinar su influencia en la red.
 - b) El análisis de redes sociales no es útil para comprender la difusión de información en una red social específica, para ello se utiliza la medida de centralidad de vector propio.
 - c) El análisis de centralidad de intermediación se utiliza para identificar usuarios que son importantes en una red social debido a su posición como “puentes” entre diferentes grupos de usuarios.
 - d) El análisis de sentimientos se utiliza para determinar la popularidad de una publicación en redes sociales sin tener en cuenta la opinión de los usuarios.
9. En la RI en el contexto de Big Data, se puede asegurar que:
- a) La indexación distribuida es una técnica obsoleta.
 - b) La indexación distribuida almacena todos los datos en un solo servidor para facilitar su acceso y búsqueda.
 - c) La indexación distribuida divide los datos en múltiples fragmentos que se almacenan en varios nodos para permitir búsquedas paralelas y mejorar la escalabilidad.
 - d) La indexación distribuida no ofrece ventajas en términos de rendimiento y escalabilidad en comparación con la indexación centralizada.
10. En un sistema de control de tráfico urbano basado en reglas, ¿cuál de las siguientes reglas sería más efectiva para manejar situaciones de congestión en una intersección?
- a) Si hay pocos vehículos en la intersección, reducir el tiempo de los semáforos en verde.
 - b) Si hay muchos vehículos en la intersección, aumentar el tiempo de los semáforos en verde.
 - c) Si hay un vehículo de emergencia en la intersección, detener todos los demás vehículos.
 - d) Si hay muchos peatones cruzando la intersección, reducir el tiempo de los semáforos en rojo.
11. La computación evolutiva:
- a) Utiliza algoritmos para buscar soluciones óptimas en grandes espacios de búsqueda, lo que la hace adecuada para problemas de optimización en la RI.
 - b) Es útil solo para problemas de clasificación de documentos y no para otras tareas de RI en general.
 - c) No es aplicable en la RI debido a la complejidad de los algoritmos evolutivos.
 - d) Solo puede manejar conjuntos de datos pequeños y no es escalable a grandes volúmenes de datos.
12. La integración de Hadoop y MapReduce en la RI trae como ventaja:
- a) La posibilidad de la extracción de información relevante y la generación de resultados significativos de grandes conjuntos de datos.
 - b) La reducción de los costos operativos a cero.
 - c) La eliminación de la necesidad de sistemas de bases de datos.
 - d) La garantía de la privacidad absoluta de los datos procesados.
13. La Web 3 se conoce como:
- a) Web de solo lectura.
 - b) Internet de las cosas.
 - c) Web de escritura-lectura.
 - d) Web semántica.
14. Los algoritmos para detectar comunidades en una red intentan:
- a) Buscar subgrafos tal que no incluyan nodos cuya ausencia desconecte al subgrafo.
 - b) Seleccionar aleatoriamente nodos de alto grado y sus vecinos.
 - c) Buscar conjuntos donde cada nodo de un mismo conjunto tenga características similares al resto de los nodos del conjunto.
 - d) Encontrar grupos donde los nodos pertenecientes a los mismos grupos son cercanos bajo cierta métrica y lejanos con respecto a los nodos de otros grupos.
15. Analizar una red permite:
- a) Evaluar la calidad del contenido de un sitio web.
 - b) Obtener predicciones exactas de eventos futuros en mercados financieros.
 - c) Detectar posibles tendencias antes de que se conviertan en tendencia.
 - d) Encontrar nodos “sensibles” o críticos para la red.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. En el contexto de la RI en redes sociales se puede afirmar que:
 - a) El análisis de redes sociales se centra exclusivamente en la cantidad de seguidores que tiene un usuario en particular para determinar su influencia en la red.
 - b) El análisis de redes sociales no es útil para comprender la difusión de información en una red social específica, para ello se utiliza la medida de centralidad de vector propio.
 - c) El análisis de centralidad de intermediación se utiliza para identificar usuarios que son importantes en una red social debido a su posición como “puentes” entre diferentes grupos de usuarios.
 - d) El análisis de sentimientos se utiliza para determinar la popularidad de una publicación en redes sociales sin tener en cuenta la opinión de los usuarios.
2. El concepto de *rank sink* en el algoritmo de PageRank representa páginas web:
 - a) Con una baja calidad de contenido y una cantidad insuficiente de enlaces salientes, lo que las hace menos relevantes en los resultados de búsqueda.
 - b) Con distribuciones uniformes de la puntuación de PageRank.
 - c) Con un alto número de enlaces salientes que no reciben enlaces entrantes, lo que puede afectar negativamente su puntuación de PageRank.
 - d) Que tienen una puntuación de PageRank más alta que otras debido a la manipulación de enlaces entrantes y salientes.
3. El componente responsable de la gestión de recursos y planificación de tareas en Hadoop es:
 - a) MapReduce.
 - b) HDFS.
 - c) Hadoop Common.
 - d) YARN.
4. En la RI en el contexto de Big Data, se puede asegurar que:
 - a) La indexación distribuida es una técnica obsoleta.
 - b) La indexación distribuida almacena todos los datos en un solo servidor para facilitar su acceso y búsqueda.
 - c) La indexación distribuida divide los datos en múltiples fragmentos que se almacenan en varios nodos para permitir búsquedas paralelas y mejorar la escalabilidad.
 - d) La indexación distribuida no ofrece ventajas en términos de rendimiento y escalabilidad en comparación con la indexación centralizada.
5. Una empresa de comercio electrónico necesita mejorar su motor de búsqueda para proporcionar resultados más relevantes a sus usuarios. Actualmente, los resultados de la búsqueda no son precisos y los usuarios a menudo encuentran dificultades para encontrar productos específicos. La empresa está considerando implementar una indexación por conceptos para mejorar la relevancia de los resultados de búsqueda. ¿Qué beneficios podría provocar este cambio?
 - a) Permite adaptarse fácilmente a cambios en el vocabulario y la terminología utilizada en las descripciones de los productos.
 - b) Facilita la visualización de productos al agruparlos por categorías o características comunes.
 - c) Mejora la precisión en la búsqueda de productos relacionados, incluso cuando no coinciden exactamente con los términos de búsqueda del usuario.
 - d) Ayuda a identificar automáticamente relaciones entre productos, lo que puede mejorar las recomendaciones personalizadas a los usuarios.
6. En el contexto del procesamiento de grandes conjuntos de datos, se puede asegurar que:
 - a) MapReduce no es adecuado para el procesamiento de datos no estructurados.
 - b) MapReduce solo puede manejar pequeñas cantidades de datos y no escala bien a grandes conjuntos de datos.
 - c) MapReduce es un enfoque para procesar datos de forma secuencial en un solo servidor para evitar problemas de concurrencia.

- d) MapReduce divide una tarea en múltiples pasos de map y reduce que se ejecutan de forma secuencial en diferentes servidores para mejorar el rendimiento y la escalabilidad.
7. ¿Qué es un 'Uniform Resource Locator (URL) Frontier' en el contexto de Web Crawling?
- Una base de datos que almacena URLs únicas identificadas como recursos en la web.
 - Una técnica para filtrar URLs irrelevantes y mejorar la eficiencia del crawling.
 - Un protocolo que define cómo se deben formatear las URLs para el crawling.
 - Una lista prioritaria de URLs que aún no han sido visitadas por el crawler.
8. Para contribuir positivamente al posicionamiento orgánico de un sitio web en los motores de búsqueda se puede:
- Crear contenido relevante y de alta calidad que satisfaga las necesidades de información de los usuarios.
 - Obtener enlaces entrantes de otros sitios web de alta autoridad y relevancia temática.
 - Incluir una densidad alta de palabras clave para asegurar que el sitio web aparezca en tantas búsquedas como sea posible.
 - Mejorar la velocidad de carga del sitio web y asegurar que sea *responsive* y fácil de usar en dispositivos móviles.
9. Sobre WordNet y su aplicación en el procesamiento del lenguaje natural, se puede afirmar que:
- Aunque WordNet es una herramienta valiosa en el NLP, su estructura no incluye información sobre la frecuencia de uso de las palabras en el lenguaje natural.
 - El diseño de WordNet facilita su integración en aplicaciones multilingües de NLP, aunque su desarrollo original se centró en el inglés.
 - Los synsets facilitan la identificación de relaciones semánticas entre palabras, como la hiperonimia y la meronimia, enriqueciendo tareas de NLP.
 - WordNet proporciona una base para la desambiguación semántica al agrupar palabras con significados similares en synsets.
10. En una empresa comercial, se tiene una red donde los nodos corresponden a los empleados y las aristas representan la frecuencia con la que colaboran en las ventas. Se busca mejorar la colaboración entre los empleados para aumentar las ventas totales, por lo que la directiva debe:
- No tener en cuenta la centralidad de cercanía de los empleados al diseñar estrategias de colaboración, ya que esto podría introducir complicaciones adicionales en el proceso de trabajo.
 - Utilizar la centralidad de cercanía para identificar los empleados menos cercanos a otros en la red y asignarles tareas individuales para evitar posibles conflictos y desacuerdos en el proceso de colaboración.
 - No considerar la centralidad de cercanía de los empleados al diseñar estrategias de colaboración, ya que todos los empleados tienen el mismo acceso a la información y los recursos.
 - Utilizar la centralidad de cercanía para identificar los empleados más cercanos a otros en la red y promover la colaboración entre ellos, facilitando así la comunicación y el intercambio de conocimientos para mejorar las ventas.
11. Existen varias estrategias consideradas como no recomendables para mejorar el SEO de un sitio web. Dentro de estas estrategias negativas, se encuentran:
- Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - Evitar el uso de las meta etiquetas y descripciones del sitio web para reflejar el contenido de la página.
 - Construir enlaces naturales de sitios web con autoridad y relevancia temática.
 - Incluir una cantidad excesiva de palabras clave irrelevantes para intentar manipular los rankings de búsqueda.
12. Sobre el algoritmo de PageRank, visto en clase, se puede afirmar que:
- Evalúa la importancia de un sitio web en función de la calidad y cantidad de enlaces entrantes que recibe de otros sitios web.
 - Asigna una puntuación baja a los sitios web que contienen muchos enlaces salientes, ya que indica una falta de relevancia.
 - Asigna una puntuación alta a los sitios web que tienen un gran número de enlaces entrantes sin tener en cuenta la calidad de esos enlaces.
 - Solo tiene en cuenta el contenido en un sitio web para determinar su relevancia en los resultados de búsqueda.
13. Considerando las dimensiones y desafíos inherentes a Big Data puede afirmarse, tomando en cuenta las características clave y las implicaciones para su procesamiento y análisis, que:

- a) Big Data se caracteriza principalmente por su pequeño volumen y uniformidad, permitiendo un procesamiento eficiente con mínimas adaptaciones de las herramientas de análisis de datos tradicionales.
 - b) Big Data no desafía la capacidad de las herramientas tradicionales de procesamiento de datos para capturar, almacenar, gestionar y analizar efectivamente la información, dada la evolución constante de las capacidades computacionales y algoritmos de optimización.
 - c) Aunque Big Data puede incluir datos estructurados, su naturaleza se expande al incorporar grandes cantidades de datos no estructurados y semiestructurados, lo que exige el uso de tecnologías especializadas en almacenamiento y procesamiento como Hadoop y sistemas de bases de datos NoSQL.
 - d) Además de su complejidad y diversidad, Big Data introduce desafíos significativos en términos de veracidad y variabilidad al requerir métodos avanzados de limpieza y validación de datos para asegurar la integridad del análisis.
14. En un sistema de control de tráfico urbano basado en reglas, ¿cuál de las siguientes reglas sería más efectiva para manejar situaciones de congestión en una intersección?
- a) Si hay pocos vehículos en la intersección, reducir el tiempo de los semáforos en verde.
 - b) Si hay muchos vehículos en la intersección, aumentar el tiempo de los semáforos en verde.
 - c) Si hay un vehículo de emergencia en la intersección, detener todos los demás vehículos.
 - d) Si hay muchos peatones cruzando la intersección, reducir el tiempo de los semáforos en rojo.
15. Una plataforma de comercio electrónico desea mejorar la experiencia del usuario al permitir una navegación más personalizada y contextualizada. Actualmente los usuarios tienen dificultades para encontrar productos relevantes debido a la gran cantidad de opciones disponibles. La empresa está interesada en implementar características de la Web 2.5 y la Web Semántica para abordar este problema. ¿Qué características podrían ayudar para ofrecer una navegación más personalizada y contextualizada?
- a) La implementación de ontologías y metadatos para enriquecer la descripción de productos y mejorar la precisión de las recomendaciones.
 - b) La optimización de la velocidad de carga del sitio web para mejorar la experiencia del usuario y reducir el abandono del carrito de compra.
 - c) La incorporación de los productos en tendencia en el mercado.
 - d) La integración de redes sociales para permitir la recomendación de productos basada en las preferencias de amigos y contactos.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. En un sistema donde el conocimiento está orientado a objetos se puede asegurar que:
 - a) Los marcos y las bandas son las estructuras utilizadas para la representación del modelo.
 - b) Las funciones de agregación dificultan poder establecer relaciones entre los objetos.
 - c) Se enfatiza la atención a la información de la cual se extrajo el conocimiento.
 - d) No es posible definir especializaciones de los objetos de la vida real dentro del sistema.
2. Para la RI, el análisis de las redes puede:
 - a) Ayudar a identificar grupos de interés.
 - b) Indicar la importancia de una entidad en la transmisión de la información.
 - c) Revelar patrones de influencia dentro de una comunidad.
 - d) Ayudar a comprender la conectividad y la accesibilidad entre las entidades.
3. La transición de la Web 1.0 a la Web 2.0 se caracterizó principalmente por:
 - a) La reducción en el uso de HTML y CSS en el desarrollo de sitios web.
 - b) La disminución de la importancia de los motores de búsqueda en la navegación web.
 - c) El cambio de páginas web estáticas a dinámicas, permitiendo la interacción del usuario y la generación de contenido.
 - d) El aumento en la velocidad de conexión a internet, que permitió una mejor calidad de las páginas web.
4. La diferencia entre la indexación por tokens y la indexación por conceptos puede definirse como:
 - a) La indexación por tokens normaliza los datos reduciéndolos a su forma básica, mientras que la indexación por conceptos utiliza un método de ordenación para organizar los términos característicos de los datos.
 - b) La indexación por tokens asigna un valor numérico a cada término de los datos, mientras que la indexación por conceptos utiliza algoritmos de encriptación para proteger la privacidad de los datos.
 - c) La indexación por tokens asigna pesos a los términos basados en su importancia relativa, mientras que la indexación por conceptos utiliza un sistema de etiquetado para asociar términos con características generales.
 - d) La indexación por tokens divide los datos en términos individuales, mientras que la indexación por conceptos agrupa los datos en categorías definidas.
5. Si una red cumple la propiedad de ser un grafo de mundo pequeño, entonces se conoce que:
 - a) La red posee pocos vértices.
 - b) La longitud media del camino entre todo par de vértices es pequeña.
 - c) El número de componentes fuertemente conexas está relacionado con la cantidad de grafos K_n presentes.
 - d) El grafo es un anillo regular de grado 5.
6. Para transformar el contenido no estructurado de las páginas web en datos estructurados, el Web Scraping:
 - a) Convierte de forma automática imágenes a texto.
 - b) Interpreta el código JavaScript en tiempo real.
 - c) Extrae información basada en patrones de HTML/CSS.
 - d) Analiza los protocolos de red.
7. La integración de Hadoop y MapReduce en la RI trae como ventaja:
 - a) La posibilidad de la extracción de información relevante y la generación de resultados significativos de grandes conjuntos de datos.
 - b) La reducción de los costos operativos a cero.
 - c) La eliminación de la necesidad de sistemas de bases de datos.
 - d) La garantía de la privacidad absoluta de los datos procesados.
8. Se puede afirmar que:
 - a) Uno de los desafíos en el procesamiento de Big Data es la capacidad de gestionar y analizar datos provenientes de diversas fuentes y en diferentes formatos de manera eficiente.

- b) MapReduce es un modelo de procesamiento distribuido utilizado para trabajar con grandes volúmenes de datos.
- c) Los SRI pueden beneficiarse de MapReduce para mejorar la RI relevante.
- d) El término “Big Data” se refiere exclusivamente al volumen de datos que una organización maneja, sin tener en cuenta la velocidad, la variedad y la veracidad de los datos.
9. Sobre el algoritmo de PageRank, visto en clase, se puede afirmar que:
- a) Evalúa la importancia de un sitio web en función de la calidad y cantidad de enlaces entrantes que recibe de otros sitios web.
- b) Asigna una puntuación baja a los sitios web que contienen muchos enlaces salientes, ya que indica una falta de relevancia.
- c) Asigna una puntuación alta a los sitios web que tienen un gran número de enlaces entrantes sin tener en cuenta la calidad de esos enlaces.
- d) Solo tiene en cuenta el contenido en un sitio web para determinar su relevancia en los resultados de búsqueda.
10. Se puede afirmar que:
- a) La cantidad de conexiones de un nodo siempre indica su influencia en la red.
- b) Todas las relaciones en una red tienen la misma importancia para el análisis.
- c) No es posible identificar subgrupos dentro de una red utilizando análisis de redes.
- d) El tamaño de una red es siempre indicativo de su efectividad en la transmisión de información.
11. La Web 1.0 se caracteriza por:
- a) Los propietarios de los sitios web proporcionan contenido de forma periódica.
- b) Uso de comunidades virtuales para popularizar los sitios web de noticias.
- c) Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
- d) Los sitios web se centran en brindar información en lugar de facilitar la colaboración o participación de los usuarios.
12. En el contexto de la representación del conocimiento basada en herencia, ¿qué caracteriza a la herencia cancelable?
- a) Las propiedades heredadas siempre se mantienen y no pueden anularse.
- b) La herencia es el resultado del razonamiento no transitivo.
- c) Las conclusiones no están determinadas y dependen del nodo de interés.
- d) No existe ambigüedad en las conclusiones obtenidas.
13. Un SRI es capaz de:
- a) No necesitar de ningún almacenamiento externo para alojar los índices de los datos.
- b) Reducir el tiempo de indexación de los datos si utiliza servidores distribuidos dentro de la red para que cada uno ejecute la indexación del mismo conjunto de datos.
- c) Generar índices invertidos de manera óptima sin considerar el contexto.
- d) Crear los índices asociados a los datos sin tener que analizar cada dato.
14. Si se tiene el conjunto de páginas interconectadas $\{A \rightarrow B, C; B \rightarrow C; C \rightarrow A\}$, entonces la página con valor más alto de PageRank es:
- a) B.
- b) C.
- c) A y B.
- d) A.
15. Existen varias estrategias consideradas como no recomendables para mejorar el SEO de un sitio web. Dentro de estas estrategias negativas, se encuentran:
- a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
- b) Evitar el uso de las meta etiquetas y descripciones del sitio web para reflejar el contenido de la página.
- c) Construir enlaces naturales de sitios web con autoridad y relevancia temática.
- d) Incluir una cantidad excesiva de palabras clave irrelevantes para intentar manipular los rankings de búsqueda.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Los Web Crawlers se enfrentan a desafíos constantes. Dentro de ellos se encuentran:
 - a) La falta de acceso a la base de datos del servidor web para extraer información actualizada.
 - b) La modificación del código y la estructura del sitio web.
 - c) La incapacidad para interpretar correctamente el lenguaje de programación utilizado en el desarrollo de los sitios web.
 - d) La dificultad para generar el contenido dinámico en tiempo real.
2. Implementar índices invertidos en un SRI asegura:
 - a) Incrementar la seguridad de los datos almacenados al dificultar el acceso directo a la información sin el índice correcto.
 - b) Reducir la cantidad de espacio de almacenamiento necesario al comprimir los datos de los documentos.
 - c) Acelerar el proceso de búsqueda al permitir búsquedas directas por contenido en lugar de por título.
 - d) Facilitar la búsqueda sobre los datos que contienen términos específicos al mantener una lista de datos para cada término único.
3. La premisa básica del algoritmo de PageRank para clasificar páginas web en los resultados de búsqueda es:
 - a) Los enlaces entrantes a una página web desde otras páginas contribuyen a su importancia y clasificación.
 - b) La antigüedad de la página web es el principal factor para determinar su clasificación.
 - c) La cantidad de visitas que recibe una página web determina su clasificación.
 - d) El contenido y la relevancia de las palabras clave en la página web determinan su posición.
4. El algoritmo Hypertext Induced Topic Selection (HITS) intenta buscar nodos especiales. Estos son conocidos como:
 - a) Sensibles.
 - b) Centrales.
 - c) Hubs.
 - d) Autoridades.
5. Se tiene un grafo donde cada nodo es un personaje de cierto libro de cuentos y la existencia de las aristas está definida si dos personajes aparecen en el mismo cuento. Se puede asegurar que:
 - a) La centralidad de grado es útil para identificar los personajes con más conexiones dentro de la red, lo que puede indicar su importancia en el libro.
 - b) La centralidad de cercanía indica el grado de conexión de cada personaje con el resto de los personajes de los cuentos del libro.
 - c) La centralidad de intermediación identifica a los personajes que actúan como conectores entre personajes de cuentos distintos.
 - d) La centralidad de cercanía ofrece una relación entre la cantidad de vecinos de un nodo con respecto a la longitud máxima de un camino dentro del grafo partiendo del nodo en cuestión.
6. En el algoritmo de Indexación Basada en Clasificación Bloqueada (BSBI), ¿cuál es el paso final para crear un índice invertido para la colección completa de los datos?
 - a) Dividir la colección de datos en bloques de tamaño fijo.
 - b) Eliminar los términos duplicados de los índices.
 - c) Indexar cada bloque de forma independiente.
 - d) Fusionar los índices invertidos de cada bloque.
7. Se puede afirmar que:
 - a) La cantidad de conexiones de un nodo siempre indica su influencia en la red.
 - b) Todas las relaciones en una red tienen la misma importancia para el análisis.
 - c) No es posible identificar subgrupos dentro de una red utilizando análisis de redes.
 - d) El tamaño de una red es siempre indicativo de su efectividad en la transmisión de información.
8. En el diseño de un videojuego de roles (RPG) en el que los personajes tienen atributos como salud, fuerza y velocidad, ¿cuál de las siguientes opciones representa mejor una implementación de la representación del conocimiento orientado a objetos?

- a) Cada personaje se representa como un objeto con propiedades como salud, fuerza y velocidad, y métodos para modificar y consultar estos valores.
 - b) Cada personaje se representa como una función que calcula sus atributos en función de su nivel y experiencia.
 - c) Cada personaje se representa como una matriz de números que almacena sus valores de atributos.
 - d) Cada personaje se representa como una lista de cadenas de texto que describen sus características físicas y habilidades.
9. Dentro del análisis de redes, la centralidad de grado mide la importancia de un nodo basándose en:
- a) La cantidad de vecinos del nodo.
 - b) El número de aristas que posee el nodo.
 - c) La cantidad de veces que aparece el nodo en el camino mínimo entre cualquier par de nodos.
 - d) El grado del nodo.
10. ¿Qué algoritmos permiten obtener información de una red?
- a) Índices de centralidad.
 - b) Detección de comunidades.
 - c) Hypertext Induced Topic Selection (HITS).
 - d) PageRank.
11. En la RI en el contexto de Big Data, se puede asegurar que:
- a) La indexación distribuida es una técnica obsoleta.
 - b) La indexación distribuida almacena todos los datos en un solo servidor para facilitar su acceso y búsqueda.
 - c) La indexación distribuida divide los datos en múltiples fragmentos que se almacenan en varios nodos para permitir búsquedas paralelas y mejorar la escalabilidad.
 - d) La indexación distribuida no ofrece ventajas en términos de rendimiento y escalabilidad en comparación con la indexación centralizada.
12. La Web 2.0 se caracteriza por:
- a) Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
 - b) Las plataformas ofrecen experiencias a la medida, permitiendo a los usuarios personalizar sus perfiles, recibir recomendaciones ajustadas al contenido y participar en filtrado colaborativo.
 - c) La existencia de sitios web dinámicos e interactivos que permiten a los usuarios participar, comentar e interactuar tanto con los creadores de contenido como con otros usuarios.
 - d) El uso de distintas tecnologías para crear experiencias web más interactivas y con mayor capacidad de respuesta.
13. Al “relajar” el concepto de clique en la detección de comunidades se intenta solucionar:
- a) La uniformidad de los nodos en términos de grado.
 - b) La necesidad de datos externos para analizar la red.
 - c) La imposibilidad de conectar nodos distantes.
 - d) El solapamiento y la complejidad computacional.
14. El componente responsable de la gestión de recursos y planificación de tareas en Hadoop es:
- a) MapReduce.
 - b) HDFS.
 - c) Hadoop Common.
 - d) YARN.
15. Existen varias estrategias consideradas como no recomendables para mejorar el SEO de un sitio web. Dentro de estas estrategias negativas, se encuentran:
- a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Evitar el uso de las meta etiquetas y descripciones del sitio web para reflejar el contenido de la página.
 - c) Construir enlaces naturales de sitios web con autoridad y relevancia temática.
 - d) Incluir una cantidad excesiva de palabras clave irrelevantes para intentar manipular los rankings de búsqueda.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Para contribuir positivamente al posicionamiento orgánico de un sitio web en los motores de búsqueda se puede:
 - a) Crear contenido relevante y de alta calidad que satisfaga las necesidades de información de los usuarios.
 - b) Obtener enlaces entrantes de otros sitios web de alta autoridad y relevancia temática.
 - c) Incluir una densidad alta de palabras clave para asegurar que el sitio web aparezca en tantas búsquedas como sea posible.
 - d) Mejorar la velocidad de carga del sitio web y asegurar que sea *responsive* y fácil de usar en dispositivos móviles.
2. En una empresa comercial, se tiene una red donde los nodos corresponden a los empleados y las aristas representan la frecuencia con la que colaboran en las ventas. Se busca mejorar la colaboración entre los empleados para aumentar las ventas totales, por lo que la directiva debe:
 - a) No tener en cuenta la centralidad de cercanía de los empleados al diseñar estrategias de colaboración, ya que esto podría introducir complicaciones adicionales en el proceso de trabajo.
 - b) Utilizar la centralidad de cercanía para identificar los empleados menos cercanos a otros en la red y asignarles tareas individuales para evitar posibles conflictos y desacuerdos en el proceso de colaboración.
 - c) No considerar la centralidad de cercanía de los empleados al diseñar estrategias de colaboración, ya que todos los empleados tienen el mismo acceso a la información y los recursos.
 - d) Utilizar la centralidad de cercanía para identificar los empleados más cercanos a otros en la red y promover la colaboración entre ellos, facilitando así la comunicación y el intercambio de conocimientos para mejorar las ventas.
3. Para transformar el contenido no estructurado de las páginas web en datos estructurados, el Web Scraping:
 - a) Convierte de forma automática imágenes a texto.
 - b) Interpreta el código JavaScript en tiempo real.
 - c) Extrae información basada en patrones de HTML/CSS.
 - d) Analiza los protocolos de red.
4. En el algoritmo de Indexación Basada en Clasificación Bloqueada (BSBI), ¿cuál es el paso final para crear un índice invertido para la colección completa de los datos?
 - a) Dividir la colección de datos en bloques de tamaño fijo.
 - b) Eliminar los términos duplicados de los índices.
 - c) Indexar cada bloque de forma independiente.
 - d) Fusionar los índices invertidos de cada bloque.
5. Un investigador necesita recopilar datos de múltiples sitios web para un estudio académico, pero se enfrenta a varios desafíos al realizar el proceso de Web Scraping de manera ética y legal. ¿Cuál de las siguientes opciones describe mejor uno de los desafíos asociados al proceso de Web Scraping?
 - a) La dificultad para encontrar herramientas de Web Scraping gratuitas y fiables.
 - b) La disponibilidad limitada de datos en línea que se pueden extraer utilizando técnicas de Web Scraping.
 - c) La necesidad de estar montado sobre un crawler que cumpla con todas las políticas.
 - d) La necesidad de comprender la estructura del sitio web y su código HTML para extraer los datos correctamente.
6. Una plataforma de comercio electrónico desea mejorar la experiencia del usuario al permitir una navegación más personalizada y contextualizada. Actualmente los usuarios tienen dificultades para encontrar productos relevantes debido a la gran cantidad de opciones disponibles. La empresa está interesada en implementar características de la Web 2.5 y la Web Semántica para abordar este problema. ¿Qué características podrían ayudar para ofrecer una navegación más personalizada y contextualizada?
 - a) La implementación de ontologías y metadatos para enriquecer la descripción de productos y mejorar la precisión de las recomendaciones.

- b) La optimización de la velocidad de carga del sitio web para mejorar la experiencia del usuario y reducir el abandono del carrito de compra.
 - c) La incorporación de los productos en tendencia en el mercado.
 - d) La integración de redes sociales para permitir la recomendación de productos basada en las preferencias de amigos y contactos.
7. Una empresa de comercio electrónico necesita mejorar su motor de búsqueda para proporcionar resultados más relevantes a sus usuarios. Actualmente, los resultados de la búsqueda no son precisos y los usuarios a menudo encuentran dificultades para encontrar productos específicos. La empresa está considerando implementar una indexación por conceptos para mejorar la relevancia de los resultados de búsqueda. ¿Qué beneficios podría provocar este cambio?
- a) Permite adaptarse fácilmente a cambios en el vocabulario y la terminología utilizada en las descripciones de los productos.
 - b) Facilita la visualización de productos al agruparlos por categorías o características comunes.
 - c) Mejora la precisión en la búsqueda de productos relacionados, incluso cuando no coinciden exactamente con los términos de búsqueda del usuario.
 - d) Ayuda a identificar automáticamente relaciones entre productos, lo que puede mejorar las recomendaciones personalizadas a los usuarios.
8. Se puede afirmar que:
- a) Uno de los desafíos en el procesamiento de Big Data es la capacidad de gestionar y analizar datos provenientes de diversas fuentes y en diferentes formatos de manera eficiente.
 - b) MapReduce es un modelo de procesamiento distribuido utilizado para trabajar con grandes volúmenes de datos.
 - c) Los SRI pueden beneficiarse de MapReduce para mejorar la RI relevante.
 - d) El término “Big Data” se refiere exclusivamente al volumen de datos que una organización maneja, sin tener en cuenta la velocidad, la variedad y la veracidad de los datos.
9. La afirmación que mejor describe la política de amabilidad en los Web Crawlers es:
- a) Los Web Crawlers son libres de recopilar datos de cualquier sitio web sin restricciones, independientemente de la cantidad de tráfico que generen.
 - b) La política de amabilidad establece pautas y reglas sobre cómo los crawlers deben interactuar con los sitios web para minimizar la carga del servidor y respetar las directivas de los administradores del sitio.
 - c) La política de amabilidad de los Web Crawlers dicta que los crawlers deben priorizar ciertos tipos de contenido sobre otros, ignorando completamente ciertas páginas web.
 - d) Los crawlers se diseñan para acceder a sitios web sin restricciones y extraer datos de manera agresiva para su indexación.
10. La premisa básica del algoritmo de PageRank para clasificar páginas web en los resultados de búsqueda es:
- a) Los enlaces entrantes a una página web desde otras páginas contribuyen a su importancia y clasificación.
 - b) La antigüedad de la página web es el principal factor para determinar su clasificación.
 - c) La cantidad de visitas que recibe una página web determina su clasificación.
 - d) El contenido y la relevancia de las palabras clave en la página web determinan su posición.
11. Una buena práctica de SEO para mejorar el posicionamiento de un sitio web en los motores de búsqueda es:
- a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Copiar contenido directamente de otros sitios web populares para aumentar la cantidad de páginas indexadas.
 - c) Llenar el contenido con palabras clave irrelevantes para aumentar la densidad de palabras clave.
 - d) Obtener enlaces de otros sitios web relevantes y de calidad que apunten al sitio.
12. El posicionamiento de un sitio web en los motores de búsqueda puede ser afectado por:
- a) Mantener una estructura de URL clara y coherente.
 - b) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - c) Contenido sin valor en el sitio web.
 - d) Obtener enlaces de sitios web irrelevantes y de baja calidad.
13. En el contexto del procesamiento de grandes conjuntos de datos, se puede asegurar que:
- a) MapReduce no es adecuado para el procesamiento de datos no estructurados.

- b) MapReduce solo puede manejar pequeñas cantidades de datos y no escala bien a grandes conjuntos de datos.
 - c) MapReduce es un enfoque para procesar datos de forma secuencial en un solo servidor para evitar problemas de concurrencia.
 - d) MapReduce divide una tarea en múltiples pasos de map y reduce que se ejecutan de forma secuencial en diferentes servidores para mejorar el rendimiento y la escalabilidad.
14. No se considera como técnica para detectar comunidades en una red:
- a) Analizar la mutualidad de los enlaces.
 - b) Usar el agrupamiento jerárquico.
 - c) Utilizar el algoritmo de K-Means.
 - d) Encontrar cliques de vértices de grado par.
15. Los Web Crawlers se enfrentan a desafíos constantes. Dentro de ellos se encuentran:
- a) La falta de acceso a la base de datos del servidor web para extraer información actualizada.
 - b) La modificación del código y la estructura del sitio web.
 - c) La incapacidad para interpretar correctamente el lenguaje de programación utilizado en el desarrollo de los sitios web.
 - d) La dificultad para generar el contenido dinámico en tiempo real.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Si una red cumple la propiedad de ser un grafo de mundo pequeño, entonces se conoce que:
 - a) La red posee pocos vértices.
 - b) La longitud media del camino entre todo par de vértices es pequeña.
 - c) El número de componentes fuertemente conexas está relacionado con la cantidad de grafos K_n presentes.
 - d) El grafo es un anillo regular de grado 5.
2. La diferencia entre la indexación por tokens y la indexación por conceptos puede definirse como:
 - a) La indexación por tokens normaliza los datos reduciéndolos a su forma básica, mientras que la indexación por conceptos utiliza un método de ordenación para organizar los términos característicos de los datos.
 - b) La indexación por tokens asigna un valor numérico a cada término de los datos, mientras que la indexación por conceptos utiliza algoritmos de encriptación para proteger la privacidad de los datos.
 - c) La indexación por tokens asigna pesos a los términos basados en su importancia relativa, mientras que la indexación por conceptos utiliza un sistema de etiquetado para asociar términos con características generales.
 - d) La indexación por tokens divide los datos en términos individuales, mientras que la indexación por conceptos agrupa los datos en categorías definidas.
3. La Web 2.0 se caracteriza por:
 - a) Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
 - b) Las plataformas ofrecen experiencias a la medida, permitiendo a los usuarios personalizar sus perfiles, recibir recomendaciones ajustadas al contenido y participar en filtrado colaborativo.
 - c) La existencia de sitios web dinámicos e interactivos que permiten a los usuarios participar, comentar e interactuar tanto con los creadores de contenido como con otros usuarios.
 - d) El uso de distintas tecnologías para crear experiencias web más interactivas y con mayor capacidad de respuesta.
4. En el contexto de la representación del conocimiento basada en herencia, ¿qué caracteriza a la herencia cancelable?
 - a) Las propiedades heredadas siempre se mantienen y no pueden anularse.
 - b) La herencia es el resultado del razonamiento no transitivo.
 - c) Las conclusiones no están determinadas y dependen del nodo de interés.
 - d) No existe ambigüedad en las conclusiones obtenidas.
5. El posicionamiento de un sitio web en los motores de búsqueda puede ser afectado por:
 - a) Mantener una estructura de URL clara y coherente.
 - b) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - c) Contenido sin valor en el sitio web.
 - d) Obtener enlaces de sitios web irrelevantes y de baja calidad.
6. Existen varias estrategias consideradas como no recomendables para mejorar el SEO de un sitio web. Dentro de estas estrategias negativas, se encuentran:
 - a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Evitar el uso de las meta etiquetas y descripciones del sitio web para reflejar el contenido de la página.
 - c) Construir enlaces naturales de sitios web con autoridad y relevancia temática.
 - d) Incluir una cantidad excesiva de palabras clave irrelevantes para intentar manipular los rankings de búsqueda.
7. Los Web Crawlers se enfrentan a desafíos constantes. Dentro de ellos se encuentran:
 - a) La falta de acceso a la base de datos del servidor web para extraer información actualizada.

- b) La modificación del código y la estructura del sitio web.
 - c) La incapacidad para interpretar correctamente el lenguaje de programación utilizado en el desarrollo de los sitios web.
 - d) La dificultad para generar el contenido dinámico en tiempo real.
8. El propósito de la política de revisitado en los Web Crawlers es:
- a) Determinar la frecuencia con la que los crawlers deben volver a visitar una URL específica para mantener la información actualizada en el índice de búsqueda.
 - b) Limitar el acceso de los crawlers a ciertos servidores luego de visitar las páginas alojadas en estos.
 - c) Establecer reglas sobre el tiempo máximo que los crawlers pueden pasar en un sitio web durante cada visita para evitar sobrecargar los servidores.
 - d) Definir la prioridad de rastreo de las URLs, determinando el orden en que los crawlers visitan y procesan cada página web.
9. En un sistema de control de tráfico urbano basado en reglas, ¿cuál de las siguientes reglas sería más efectiva para manejar situaciones de congestión en una intersección?
- a) Si hay pocos vehículos en la intersección, reducir el tiempo de los semáforos en verde.
 - b) Si hay muchos vehículos en la intersección, aumentar el tiempo de los semáforos en verde.
 - c) Si hay un vehículo de emergencia en la intersección, detener todos los demás vehículos.
 - d) Si hay muchos peatones cruzando la intersección, reducir el tiempo de los semáforos en rojo.
10. La integración de Hadoop y MapReduce en la RI trae como ventaja:
- a) La posibilidad de la extracción de información relevante y la generación de resultados significativos de grandes conjuntos de datos.
 - b) La reducción de los costos operativos a cero.
 - c) La eliminación de la necesidad de sistemas de bases de datos.
 - d) La garantía de la privacidad absoluta de los datos procesados.
11. En un sistema donde el conocimiento está definido a partir de reglas se puede asegurar que:
- a) El nivel de especificidad de las reglas está limitado.
- b) La representación de conocimiento está basada en la lógica proposicional.
 - c) El razonamiento se activa en cadena hacia delante.
 - d) El orden en que se definen las reglas no altera el razonamiento del sistema.
12. En un sistema donde el conocimiento está orientado a objetos se puede asegurar que:
- a) Los marcos y las bandas son las estructuras utilizadas para la representación del modelo.
 - b) Las funciones de agregación dificultan poder establecer relaciones entre los objetos.
 - c) Se enfatiza la atención a la información de la cual se extrajo el conocimiento.
 - d) No es posible definir especializaciones de los objetos de la vida real dentro del sistema.
13. La afirmación que mejor describe la política de amabilidad en los Web Crawlers es:
- a) Los Web Crawlers son libres de recopilar datos de cualquier sitio web sin restricciones, independientemente de la cantidad de tráfico que generen.
 - b) La política de amabilidad establece pautas y reglas sobre cómo los crawlers deben interactuar con los sitios web para minimizar la carga del servidor y respetar las directivas de los administradores del sitio.
 - c) La política de amabilidad de los Web Crawlers dicta que los crawlers deben priorizar ciertos tipos de contenido sobre otros, ignorando completamente ciertas páginas web.
 - d) Los crawlers se diseñan para acceder a sitios web sin restricciones y extraer datos de manera agresiva para su indexación.
14. Considerando las dimensiones y desafíos inherentes a Big Data puede afirmarse, tomando en cuenta las características clave y las implicaciones para su procesamiento y análisis, que:
- a) Big Data se caracteriza principalmente por su pequeño volumen y uniformidad, permitiendo un procesamiento eficiente con mínimas adaptaciones de las herramientas de análisis de datos tradicionales.
 - b) Big Data no desafía la capacidad de las herramientas tradicionales de procesamiento de datos para capturar, almacenar, gestionar y analizar efectivamente la información, dada la evolución constante de las capacidades computacionales y algoritmos de optimización.

- c) Aunque Big Data puede incluir datos estructurados, su naturaleza se expande al incorporar grandes cantidades de datos no estructurados y semiestructurados, lo que exige el uso de tecnologías especializadas en almacenamiento y procesamiento como Hadoop y sistemas de bases de datos NoSQL.
 - d) Además de su complejidad y diversidad, Big Data introduce desafíos significativos en términos de veracidad y variabilidad al requerir métodos avanzados de limpieza y validación de datos para asegurar la integridad del análisis.
15. Se puede afirmar que:
- a) La cantidad de conexiones de un nodo siempre indica su influencia en la red.
 - b) Todas las relaciones en una red tienen la misma importancia para el análisis.
 - c) No es posible identificar subgrupos dentro de una red utilizando análisis de redes.
 - d) El tamaño de una red es siempre indicativo de su efectividad en la transmisión de información.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. En la RI en el contexto de Big Data, se puede asegurar que:
 - a) El almacenamiento distribuido es una técnica obsoleta.
 - b) El almacenamiento distribuido centraliza todos los datos en un único servidor para facilitar su gestión y mantenimiento.
 - c) El almacenamiento distribuido reparte los datos en múltiples servidores para mejorar la disponibilidad y la redundancia del sistema.
 - d) El almacenamiento distribuido ofrece ventajas significativas en términos de escalabilidad y rendimiento en comparación con el almacenamiento centralizado.
2. La Web 2.0 se caracteriza por:
 - a) Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
 - b) Las plataformas ofrecen experiencias a la medida, permitiendo a los usuarios personalizar sus perfiles, recibir recomendaciones ajustadas al contenido y participar en filtrado colaborativo.
 - c) La existencia de sitios web dinámicos e interactivos que permiten a los usuarios participar, comentar e interactuar tanto con los creadores de contenido como con otros usuarios.
 - d) El uso de distintas tecnologías para crear experiencias web más interactivas y con mayor capacidad de respuesta.
3. En el contexto de la RI en redes sociales se puede afirmar que:
 - a) El análisis de redes sociales se centra exclusivamente en la cantidad de seguidores que tiene un usuario en particular para determinar su influencia en la red.
 - b) El análisis de redes sociales no es útil para comprender la difusión de información en una red social específica, para ello se utiliza la medida de centralidad de vector propio.
 - c) El análisis de centralidad de intermediación se utiliza para identificar usuarios que son importantes en una red social debido a su posición como “puentes” entre diferentes grupos de usuarios.
 - d) El análisis de sentimientos se utiliza para determinar la popularidad de una publicación en redes sociales sin tener en cuenta la opinión de los usuarios.
4. Sobre WordNet y su aplicación en el procesamiento del lenguaje natural, se puede afirmar que:
 - a) Aunque WordNet es una herramienta valiosa en el NLP, su estructura no incluye información sobre la frecuencia de uso de las palabras en el lenguaje natural.
 - b) El diseño de WordNet facilita su integración en aplicaciones multilingües de NLP, aunque su desarrollo original se centró en el inglés.
 - c) Los synsets facilitan la identificación de relaciones semánticas entre palabras, como la hiperonimia y la meronimia, enriqueciendo tareas de NLP.
 - d) WordNet proporciona una base para la desambiguación semántica al agrupar palabras con significados similares en synsets.
5. Una plataforma de comercio electrónico desea mejorar la experiencia del usuario al permitir una navegación más personalizada y contextualizada. Actualmente los usuarios tienen dificultades para encontrar productos relevantes debido a la gran cantidad de opciones disponibles. La empresa está interesada en implementar características de la Web 2.5 y la Web Semántica para abordar este problema. ¿Qué características podrían ayudar para ofrecer una navegación más personalizada y contextualizada?
 - a) La implementación de ontologías y metadatos para enriquecer la descripción de productos y mejorar la precisión de las recomendaciones.
 - b) La optimización de la velocidad de carga del sitio web para mejorar la experiencia del usuario y reducir el abandono del carrito de compra.
 - c) La incorporación de los productos en tendencia en el mercado.
 - d) La integración de redes sociales para permitir la recomendación de productos basada en las preferencias de amigos y contactos.

6. El propósito de la política de revisitado en los Web Crawlers es:
 - a) Determinar la frecuencia con la que los crawlers deben volver a visitar una URL específica para mantener la información actualizada en el índice de búsqueda.
 - b) Limitar el acceso de los crawlers a ciertos servidores luego de visitar las páginas alojadas en estos.
 - c) Establecer reglas sobre el tiempo máximo que los crawlers pueden pasar en un sitio web durante cada visita para evitar sobrecargar los servidores.
 - d) Definir la prioridad de rastreo de las URLs, determinando el orden en que los crawlers visitan y procesan cada página web.
7. ¿Qué es un 'Uniform Resource Locator (URL) Frontier' en el contexto de Web Crawling?
 - a) Una base de datos que almacena URLs únicas identificadas como recursos en la web.
 - b) Una técnica para filtrar URLs irrelevantes y mejorar la eficiencia del crawling.
 - c) Un protocolo que define cómo se deben formatear las URLs para el crawling.
 - d) Una lista prioritaria de URLs que aún no han sido visitadas por el crawler.
8. La transición de la Web 1.0 a la Web 2.0 se caracterizó principalmente por:
 - a) La reducción en el uso de HTML y CSS en el desarrollo de sitios web.
 - b) La disminución de la importancia de los motores de búsqueda en la navegación web.
 - c) El cambio de páginas web estáticas a dinámicas, permitiendo la interacción del usuario y la generación de contenido.
 - d) El aumento en la velocidad de conexión a internet, que permitió una mejor calidad de las páginas web.
9. El algoritmo Hypertext Induced Topic Selection (HITS) intenta buscar nodos especiales. Estos son conocidos como:
 - a) Sensibles.
 - b) Centrales.
 - c) Hubs.
 - d) Autoridades.
10. La computación evolutiva:
 - a) Utiliza algoritmos para buscar soluciones óptimas en grandes espacios de búsqueda, lo que la hace adecuada para problemas de optimización en la RI.
 - b) Es útil solo para problemas de clasificación de documentos y no para otras tareas de RI en general.
 - c) No es aplicable en la RI debido a la complejidad de los algoritmos evolutivos.
 - d) Solo puede manejar conjuntos de datos pequeños y no es escalable a grandes volúmenes de datos.
11. Al "relajar" el concepto de clique en la detección de comunidades se intenta solucionar:
 - a) La uniformidad de los nodos en términos de grado.
 - b) La necesidad de datos externos para analizar la red.
 - c) La imposibilidad de conectar nodos distantes.
 - d) El solapamiento y la complejidad computacional.
12. En el diseño de un videojuego de roles (RPG) en el que los personajes tienen atributos como salud, fuerza y velocidad, ¿cuál de las siguientes opciones representa mejor una implementación de la representación del conocimiento orientado a objetos?
 - a) Cada personaje se representa como un objeto con propiedades como salud, fuerza y velocidad, y métodos para modificar y consultar estos valores.
 - b) Cada personaje se representa como una función que calcula sus atributos en función de su nivel y experiencia.
 - c) Cada personaje se representa como una matriz de números que almacena sus valores de atributos.
 - d) Cada personaje se representa como una lista de cadenas de texto que describen sus características físicas y habilidades.
13. El uso de Hadoop y MapReduce en el contexto de la RI tiene como objetivo:
 - a) Facilitar el análisis en tiempo real de datos de redes sociales.
 - b) Automatizar el mantenimiento de sistemas de bases de datos.
 - c) Mejorar la eficiencia energética en centros de datos.
 - d) Procesar y analizar grandes conjuntos de datos para la RI.
14. ¿Cuál de las siguientes opciones describe mejor la diferencia clave entre Web Crawling y Web Scraping?
 - a) Web Scraping implica el análisis de la estructura y el contenido de las páginas web para extraer datos, mientras que Web Crawling se refiere a la descarga y almacenamiento de páginas web completas.

- b) Web Scraping es más eficaz para rastrear e indexar contenido web para motores de búsqueda, mientras que Web Crawling se utiliza principalmente para la extracción de datos en proyectos de investigación.
 - c) Web Crawling se centra en la exploración y recopilación de enlaces de múltiples sitios web, mientras que Web Scraping se enfoca en la extracción específica de datos de páginas web individuales.
 - d) Web Crawling se realiza utilizando herramientas de automatización como Selenium WebDriver, mientras que Web Scraping se lleva a cabo mediante el análisis de HTML y CSS.
15. El posicionamiento de un sitio web en los motores de búsqueda puede ser afectado por:
- a) Mantener una estructura de URL clara y coherente.
 - b) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - c) Contenido sin valor en el sitio web.
 - d) Obtener enlaces de sitios web irrelevantes y de baja calidad.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. ¿Qué algoritmos permiten obtener información de una red?
 - a) Índices de centralidad.
 - b) Detección de comunidades.
 - c) Hypertext Induced Topic Selection (HITS).
 - d) PageRank.
2. El componente responsable de la gestión de recursos y planificación de tareas en Hadoop es:
 - a) MapReduce.
 - b) HDFS.
 - c) Hadoop Common.
 - d) YARN.
3. En un sistema donde el conocimiento está orientado a objetos se puede asegurar que:
 - a) Los marcos y las bandas son las estructuras utilizadas para la representación del modelo.
 - b) Las funciones de agregación dificultan poder establecer relaciones entre los objetos.
 - c) Se enfatiza la atención a la información de la cual se extrajo el conocimiento.
 - d) No es posible definir especializaciones de los objetos de la vida real dentro del sistema.
4. La Web 1.0 se caracteriza por:
 - a) Los propietarios de los sitios web proporcionan contenido de forma periódica.
 - b) Uso de comunidades virtuales para popularizar los sitios web de noticias.
 - c) Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
 - d) Los sitios web se centran en brindar información en lugar de facilitar la colaboración o participación de los usuarios.
5. El uso de Hadoop y MapReduce en el contexto de la RI tiene como objetivo:
 - a) Facilitar el análisis en tiempo real de datos de redes sociales.
 - b) Automatizar el mantenimiento de sistemas de bases de datos.
 - c) Mejorar la eficiencia energética en centros de datos.
 - d) Procesar y analizar grandes conjuntos de datos para la RI.
6. En el contexto de la RI en redes, se puede afirmar que:
 - a) La detección de comunidades ayuda a identificar grupos de nodos altamente conectados entre sí, lo que puede ser útil para comprender la estructura y el contenido de la red.
 - b) La detección de comunidades en una red siempre produce resultados objetivos y consistentes independientemente del algoritmo utilizado.
 - c) La detección de comunidades solo se aplica a redes pequeñas y simples, no a redes grandes y complejas.
 - d) La detección de comunidades no es relevante para la RI en redes, ya que se centra únicamente en la estructura de la red sin considerar el contenido.
7. Un sistema cuenta con la siguiente información:
 - Juan pertenece a la marina.
 - Juan es capellán (sacerdote encargado de una tarea específica fuera de la parroquia).
 - Los infantes de la marina suelen ser bebedores de cerveza.
 - Un capellán no suele ser bebedor de cerveza.
 - Un bebedor de cerveza suele tener sobrepeso.
 - Por lo general, un infante de la marina está en buena condición física.

Si se representa la información en una red de herencia se puede concluir que:

 - a) Existen dos razonamientos cancelables.
 - b) Juan tiene sobrepeso.
 - c) No se puede asegurar que Juan sea capellán.
 - d) No se puede asegurar que Juan esté en buena condición física.
8. En un sistema de control de tráfico urbano basado en reglas, ¿cuál de las siguientes reglas sería más efectiva para manejar situaciones de congestión en una intersección?

- a) Si hay pocos vehículos en la intersección, reducir el tiempo de los semáforos en verde.
 - b) Si hay muchos vehículos en la intersección, aumentar el tiempo de los semáforos en verde.
 - c) Si hay un vehículo de emergencia en la intersección, detener todos los demás vehículos.
 - d) Si hay muchos peatones cruzando la intersección, reducir el tiempo de los semáforos en rojo.
9. El algoritmo de PageRank converge si:
- a) Se define un factor de normalización en la fórmula de la función.
 - b) La norma de la diferencia entre los vectores es menor a un umbral predefinido.
 - c) Finaliza la ejecución cuando el número de iteraciones excede un máximo de iteraciones previamente definido.
 - d) El algoritmo no se implementa de forma iterativa.
10. La afirmación que mejor describe la política de amabilidad en los Web Crawlers es:
- a) Los Web Crawlers son libres de recopilar datos de cualquier sitio web sin restricciones, independientemente de la cantidad de tráfico que generen.
 - b) La política de amabilidad establece pautas y reglas sobre cómo los crawlers deben interactuar con los sitios web para minimizar la carga del servidor y respetar las directivas de los administradores del sitio.
 - c) La política de amabilidad de los Web Crawlers dicta que los crawlers deben priorizar ciertos tipos de contenido sobre otros, ignorando completamente ciertas páginas web.
 - d) Los crawlers se diseñan para acceder a sitios web sin restricciones y extraer datos de manera agresiva para su indexación.
11. En un SRI la indexación:
- a) Mejora la experiencia del usuario.
 - b) Consiste en asociar un identificador único a cada dato almacenado en el sistema.
 - c) Facilita la RI relevante.
 - d) Permite la organización y la categorización de la información.
12. Al realizar la optimización de contenido para SEO debe considerarse:
- a) Utilizar etiquetas de título y meta descripciones únicas y relevantes para cada página.
 - b) Seleccionar las palabras clave al azar.
 - c) Crear contenido valioso y original que satisfaga las necesidades de los usuarios.
 - d) Incluir palabras clave de manera excesiva en el contenido para mejorar el posicionamiento.
13. Al “relajar” el concepto de clique en la detección de comunidades se intenta solucionar:
- a) La uniformidad de los nodos en términos de grado.
 - b) La necesidad de datos externos para analizar la red.
 - c) La imposibilidad de conectar nodos distantes.
 - d) El solapamiento y la complejidad computacional.
14. La transición de la Web 1.0 a la Web 2.0 se caracterizó principalmente por:
- a) La reducción en el uso de HTML y CSS en el desarrollo de sitios web.
 - b) La disminución de la importancia de los motores de búsqueda en la navegación web.
 - c) El cambio de páginas web estáticas a dinámicas, permitiendo la interacción del usuario y la generación de contenido.
 - d) El aumento en la velocidad de conexión a internet, que permitió una mejor calidad de las páginas web.
15. En el contexto de la RI en redes sociales se puede afirmar que:
- a) El análisis de redes sociales se centra exclusivamente en la cantidad de seguidores que tiene un usuario en particular para determinar su influencia en la red.
 - b) El análisis de redes sociales no es útil para comprender la difusión de información en una red social específica, para ello se utiliza la medida de centralidad de vector propio.
 - c) El análisis de centralidad de intermediación se utiliza para identificar usuarios que son importantes en una red social debido a su posición como “puentes” entre diferentes grupos de usuarios.
 - d) El análisis de sentimientos se utiliza para determinar la popularidad de una publicación en redes sociales sin tener en cuenta la opinión de los usuarios.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Sobre el algoritmo de PageRank, visto en clase, se puede afirmar que:
 - a) Evalúa la importancia de un sitio web en función de la calidad y cantidad de enlaces entrantes que recibe de otros sitios web.
 - b) Asigna una puntuación baja a los sitios web que contienen muchos enlaces salientes, ya que indica una falta de relevancia.
 - c) Asigna una puntuación alta a los sitios web que tienen un gran número de enlaces entrantes sin tener en cuenta la calidad de esos enlaces.
 - d) Solo tiene en cuenta el contenido en un sitio web para determinar su relevancia en los resultados de búsqueda.
2. La computación evolutiva:
 - a) Utiliza algoritmos para buscar soluciones óptimas en grandes espacios de búsqueda, lo que la hace adecuada para problemas de optimización en la RI.
 - b) Es útil solo para problemas de clasificación de documentos y no para otras tareas de RI en general.
 - c) No es aplicable en la RI debido a la complejidad de los algoritmos evolutivos.
 - d) Solo puede manejar conjuntos de datos pequeños y no es escalable a grandes volúmenes de datos.
3. Sobre WordNet y su aplicación en el procesamiento del lenguaje natural, se puede afirmar que:
 - a) Aunque WordNet es una herramienta valiosa en el NLP, su estructura no incluye información sobre la frecuencia de uso de las palabras en el lenguaje natural.
 - b) El diseño de WordNet facilita su integración en aplicaciones multilingües de NLP, aunque su desarrollo original se centró en el inglés.
 - c) Los synsets facilitan la identificación de relaciones semánticas entre palabras, como la hiperonimia y la meronimia, enriqueciendo tareas de NLP.
 - d) WordNet proporciona una base para la desambiguación semántica al agrupar palabras con significados similares en synsets.
4. Para transformar el contenido no estructurado de las páginas web en datos estructurados, el Web Scraping:
 - a) Convierte de forma automática imágenes a texto.
 - b) Interpreta el código JavaScript en tiempo real.
 - c) Extrae información basada en patrones de HTML/CSS.
 - d) Analiza los protocolos de red.
5. Se tiene un grafo donde cada nodo es un personaje de cierto libro de cuentos y la existencia de las aristas está definida si dos personajes aparecen en el mismo cuento. Se puede asegurar que:
 - a) La centralidad de grado es útil para identificar los personajes con más conexiones dentro de la red, lo que puede indicar su importancia en el libro.
 - b) La centralidad de cercanía indica el grado de conexión de cada personaje con el resto de los personajes de los cuentos del libro.
 - c) La centralidad de intermediación identifica a los personajes que actúan como conectores entre personajes de cuentos distintos.
 - d) La centralidad de cercanía ofrece una relación entre la cantidad de vecinos de un nodo con respecto a la longitud máxima de un camino dentro del grafo partiendo del nodo en cuestión.
6. La diferencia entre la indexación por tokens y la indexación por conceptos puede definirse como:
 - a) La indexación por tokens normaliza los datos reduciéndolos a su forma básica, mientras que la indexación por conceptos utiliza un método de ordenación para organizar los términos característicos de los datos.
 - b) La indexación por tokens asigna un valor numérico a cada término de los datos, mientras que la indexación por conceptos utiliza algoritmos de encriptación para proteger la privacidad de los datos.
 - c) La indexación por tokens asigna pesos a los términos basados en su importancia relativa, mientras que la indexación por conceptos utiliza un sistema de etiquetado para asociar términos con características generales.

- d) La indexación por tokens divide los datos en términos individuales, mientras que la indexación por conceptos agrupa los datos en categorías definidas.
7. La política de ordenación de URLs en los Web Crawlers tiene como aspecto fundamental:
- Limitar el acceso de los crawlers a ciertas secciones de un sitio web, evitando el rastreo de URLs consideradas menos importantes o sensibles.
 - Establecer la prioridad de rastreo de las URLs, determinando el orden en que los crawlers visitan y procesan cada página web.
 - Definir la estructura de la URL de destino, asegurando que estén ordenadas alfabéticamente para facilitar la navegación y la indexación.
 - Determinar la forma en que los crawlers asignan un valor de relevancia a cada URL para clasificarlas en el índice de búsqueda.
8. Un sistema cuenta con la siguiente información:
- Juan pertenece a la marina.
 - Juan es capellán (sacerdote encargado de una tarea específica fuera de la parroquia).
 - Los infantes de la marina suelen ser bebedores de cerveza.
 - Un capellán no suele ser bebedor de cerveza.
 - Un bebedor de cerveza suele tener sobrepeso.
 - Por lo general, un infante de la marina está en buena condición física.
- Si se representa la información en una red de herencia se puede concluir que:
- Existen dos razonamientos cancelables.
 - Juan tiene sobrepeso.
 - No se puede asegurar que Juan sea capellán.
 - No se puede asegurar que Juan esté en buena condición física.
9. El conjunto frontera de URLs en un Web Crawler:
- Delimita los sitios web que visitará en futuras iteraciones del proceso.
 - Indica las secciones que no pueden ser visitadas de cada sitio web.
 - Almacena hipervínculos que pertenecen al mismo dominio del conjunto semilla de URLs con que inició el crawler.
 - Es similar a un conjunto de URLs que esperan ser visitadas.
10. No se considera como técnica para detectar comunidades en una red:
- Analizar la mutualidad de los enlaces.
 - Usar el agrupamiento jerárquico.
 - Utilizar el algoritmo de K-Means.
 - Encontrar cliques de vértices de grado par.
11. La integración de Hadoop y MapReduce en la RI trae como ventaja:
- La posibilidad de la extracción de información relevante y la generación de resultados significativos de grandes conjuntos de datos.
 - La reducción de los costos operativos a cero.
 - La eliminación de la necesidad de sistemas de bases de datos.
 - La garantía de la privacidad absoluta de los datos procesados.
12. Considerando las dimensiones y desafíos inherentes a Big Data puede afirmarse, tomando en cuenta las características clave y las implicaciones para su procesamiento y análisis, que:
- Big Data se caracteriza principalmente por su pequeño volumen y uniformidad, permitiendo un procesamiento eficiente con mínimas adaptaciones de las herramientas de análisis de datos tradicionales.
 - Big Data no desafía la capacidad de las herramientas tradicionales de procesamiento de datos para capturar, almacenar, gestionar y analizar efectivamente la información, dada la evolución constante de las capacidades computacionales y algoritmos de optimización.
 - Aunque Big Data puede incluir datos estructurados, su naturaleza se expande al incorporar grandes cantidades de datos no estructurados y semiestructurados, lo que exige el uso de tecnologías especializadas en almacenamiento y procesamiento como Hadoop y sistemas de bases de datos NoSQL.
 - Además de su complejidad y diversidad, Big Data introduce desafíos significativos en términos de veracidad y variabilidad al requerir métodos avanzados de limpieza y validación de datos para asegurar la integridad del análisis.
13. Una buena práctica de SEO para mejorar el posicionamiento de un sitio web en los motores de búsqueda es:
- Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - Copiar contenido directamente de otros sitios web populares para aumentar la cantidad de páginas indexadas.
 - Llenar el contenido con palabras clave irrelevantes para aumentar la densidad de palabras clave.

- d) Obtener enlaces de otros sitios web relevantes y de calidad que apunten al sitio.
14. ¿Cuál de las siguientes opciones describe mejor la diferencia clave entre Web Crawling y Web Scraping?
- a) Web Scraping implica el análisis de la estructura y el contenido de las páginas web para extraer datos, mientras que Web Crawling se refiere a la descarga y almacenamiento de páginas web completas.
 - b) Web Scraping es más eficaz para rastrear e indexar contenido web para motores de búsqueda, mientras que Web Crawling se utiliza principalmente para la extracción de datos en proyectos de investigación.
 - c) Web Crawling se centra en la exploración y recopilación de enlaces de múltiples sitios web, mientras que Web Scraping se enfoca en la extracción específica de datos de páginas web individuales.
- d) Web Crawling se realiza utilizando herramientas de automatización como Selenium WebDriver, mientras que Web Scraping se lleva a cabo mediante el análisis de HTML y CSS.
15. Considerando las prácticas éticas y legales en el Web Scraping, se puede asegurar que:
- a) Web Scraping sobre datos personales sin consentimiento es generalmente aceptado si los datos se utilizan con fines de investigación.
 - b) Es importante revisar y respetar los términos de servicio del sitio web, así como las leyes aplicables de protección de datos y derechos de autor, antes de realizar Web Scraping.
 - c) Web Scraping requiere considerar las políticas de `robots.txt` del sitio web objetivo.
 - d) La extracción de datos mediante Web Scraping siempre es legal, independientemente de las leyes locales sobre derechos de autor y privacidad de datos definidos en los sitios web.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. El componente responsable de la gestión de recursos y planificación de tareas en Hadoop es:
 - a) MapReduce.
 - b) HDFS.
 - c) Hadoop Common.
 - d) YARN.
2. En el contexto del procesamiento de grandes conjuntos de datos, se puede asegurar que:
 - a) MapReduce no es adecuado para el procesamiento de datos no estructurados.
 - b) MapReduce solo puede manejar pequeñas cantidades de datos y no escala bien a grandes conjuntos de datos.
 - c) MapReduce es un enfoque para procesar datos de forma secuencial en un solo servidor para evitar problemas de concurrencia.
 - d) MapReduce divide una tarea en múltiples pasos de map y reduce que se ejecutan de forma secuencial en diferentes servidores para mejorar el rendimiento y la escalabilidad.
3. En el contexto de la RI en redes sociales se puede afirmar que:
 - a) El análisis de redes sociales se centra exclusivamente en la cantidad de seguidores que tiene un usuario en particular para determinar su influencia en la red.
 - b) El análisis de redes sociales no es útil para comprender la difusión de información en una red social específica, para ello se utiliza la medida de centralidad de vector propio.
 - c) El análisis de centralidad de intermediación se utiliza para identificar usuarios que son importantes en una red social debido a su posición como “puentes” entre diferentes grupos de usuarios.
 - d) El análisis de sentimientos se utiliza para determinar la popularidad de una publicación en redes sociales sin tener en cuenta la opinión de los usuarios.
4. La Web 2.0 se caracteriza por:
 - a) Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
 - b) Las plataformas ofrecen experiencias a la medida, permitiendo a los usuarios personalizar sus perfiles, recibir recomendaciones ajustadas al contenido y participar en filtrado colaborativo.
 - c) La existencia de sitios web dinámicos e interactivos que permiten a los usuarios participar, comentar e interactuar tanto con los creadores de contenido como con otros usuarios.
 - d) El uso de distintas tecnologías para crear experiencias web más interactivas y con mayor capacidad de respuesta.
5. Si una red cumple la propiedad de ser un grafo de mundo pequeño, entonces se conoce que:
 - a) La red posee pocos vértices.
 - b) La longitud media del camino entre todo par de vértices es pequeña.
 - c) El número de componentes fuertemente conexas está relacionado con la cantidad de grafos K_n presentes.
 - d) El grafo es un anillo regular de grado 5.
6. La centralidad de intermediación de un nodo indica:
 - a) La frecuencia con la que un nodo actúa como puente en el camino más corto entre otros dos nodos.
 - b) El número total de conexiones entrantes y salientes.
 - c) La resistencia del nodo a fallos.
 - d) La cantidad de nodos vecinos directos.
7. Al “relajar” el concepto de clique en la detección de comunidades se intenta solucionar:
 - a) La uniformidad de los nodos en términos de grado.
 - b) La necesidad de datos externos para analizar la red.
 - c) La imposibilidad de conectar nodos distantes.
 - d) El solapamiento y la complejidad computacional.

8. En un sistema de control de tráfico urbano basado en reglas, ¿cuál de las siguientes reglas sería más efectiva para manejar situaciones de congestión en una intersección?
 - a) Si hay pocos vehículos en la intersección, reducir el tiempo de los semáforos en verde.
 - b) Si hay muchos vehículos en la intersección, aumentar el tiempo de los semáforos en verde.
 - c) Si hay un vehículo de emergencia en la intersección, detener todos los demás vehículos.
 - d) Si hay muchos peatones cruzando la intersección, reducir el tiempo de los semáforos en rojo.
9. En un sistema donde el conocimiento está definido a partir de reglas se puede asegurar que:
 - a) El nivel de especificidad de las reglas está limitado.
 - b) La representación de conocimiento está basada en la lógica proposicional.
 - c) El razonamiento se activa en cadena hacia delante.
 - d) El orden en que se definen las reglas no altera el razonamiento del sistema.
10. Sobre el algoritmo de PageRank, visto en clase, se puede afirmar que:
 - a) Evalúa la importancia de un sitio web en función de la calidad y cantidad de enlaces entrantes que recibe de otros sitios web.
 - b) Asigna una puntuación baja a los sitios web que contienen muchos enlaces salientes, ya que indica una falta de relevancia.
 - c) Asigna una puntuación alta a los sitios web que tienen un gran número de enlaces entrantes sin tener en cuenta la calidad de esos enlaces.
 - d) Solo tiene en cuenta el contenido en un sitio web para determinar su relevancia en los resultados de búsqueda.
11. ¿Qué algoritmos permiten obtener información de una red?
 - a) Índices de centralidad.
 - b) Detección de comunidades.
 - c) Hypertext Induced Topic Selection (HITS).
 - d) PageRank.
12. ¿Qué es un 'Uniform Resource Locator (URL) Frontier' en el contexto de Web Crawling?
 - a) Una base de datos que almacena URLs únicas identificadas como recursos en la web.
 - b) Una técnica para filtrar URLs irrelevantes y mejorar la eficiencia del crawling.
 - c) Un protocolo que define cómo se deben formatear las URLs para el crawling.
 - d) Una lista prioritaria de URLs que aún no han sido visitadas por el crawler.
13. Se puede afirmar que:
 - a) La cantidad de conexiones de un nodo siempre indica su influencia en la red.
 - b) Todas las relaciones en una red tienen la misma importancia para el análisis.
 - c) No es posible identificar subgrupos dentro de una red utilizando análisis de redes.
 - d) El tamaño de una red es siempre indicativo de su efectividad en la transmisión de información.
14. ¿Qué estrategia utilizan los Web Crawlers para asegurar un rastreo eficiente y respetuoso de los recursos de los sitios web?
 - a) Extraer únicamente contenido multimedia para reducir la carga en los servidores web.
 - b) Visitar y rastrear todos los enlaces de una página web simultáneamente.
 - c) Ignorar completamente el archivo `robots.txt` de los sitios web.
 - d) Seguir las directrices del archivo `robots.txt` y aplicar un retraso entre las solicitudes.
15. Existen varias estrategias consideradas como no recomendables para mejorar el SEO de un sitio web. Dentro de estas estrategias negativas, se encuentran:
 - a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Evitar el uso de las meta etiquetas y descripciones del sitio web para reflejar el contenido de la página.
 - c) Construir enlaces naturales de sitios web con autoridad y relevancia temática.
 - d) Incluir una cantidad excesiva de palabras clave irrelevantes para intentar manipular los rankings de búsqueda.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. ¿Qué es un 'Uniform Resource Locator (URL) Frontier' en el contexto de Web Crawling?
 - a) Una base de datos que almacena URLs únicas identificadas como recursos en la web.
 - b) Una técnica para filtrar URLs irrelevantes y mejorar la eficiencia del crawling.
 - c) Un protocolo que define cómo se deben formatear las URLs para el crawling.
 - d) Una lista prioritaria de URLs que aún no han sido visitadas por el crawler.
2. En el algoritmo de Indexación Basada en Clasificación Bloqueada (BSBI), ¿cuál es el paso final para crear un índice invertido para la colección completa de los datos?
 - a) Dividir la colección de datos en bloques de tamaño fijo.
 - b) Eliminar los términos duplicados de los índices.
 - c) Indexar cada bloque de forma independiente.
 - d) Fusionar los índices invertidos de cada bloque.
3. Un sistema cuenta con la siguiente información:
 - Juan pertenece a la marina.
 - Juan es capellán (sacerdote encargado de una tarea específica fuera de la parroquia).
 - Los infantes de la marina suelen ser bebedores de cerveza.
 - Un capellán no suele ser bebedor de cerveza.
 - Un bebedor de cerveza suele tener sobrepeso.
 - Por lo general, un infante de la marina está en buena condición física.

Si se representa la información en una red de herencia se puede concluir que:

 - a) Existen dos razonamientos cancelables.
 - b) Juan tiene sobrepeso.
 - c) No se puede asegurar que Juan sea capellán.
 - d) No se puede asegurar que Juan esté en buena condición física.
4. Considerando las dimensiones y desafíos inherentes a Big Data puede afirmarse, tomando en cuenta las características clave y las implicaciones para su procesamiento y análisis, que:
 - a) Big Data se caracteriza principalmente por su pequeño volumen y uniformidad, permitiendo un procesamiento eficiente con mínimas adaptaciones de las herramientas de análisis de datos tradicionales.
 - b) Big Data no desafía la capacidad de las herramientas tradicionales de procesamiento de datos para capturar, almacenar, gestionar y analizar efectivamente la información, dada la evolución constante de las capacidades computacionales y algoritmos de optimización.
 - c) Aunque Big Data puede incluir datos estructurados, su naturaleza se expande al incorporar grandes cantidades de datos no estructurados y semiestructurados, lo que exige el uso de tecnologías especializadas en almacenamiento y procesamiento como Hadoop y sistemas de bases de datos NoSQL.
 - d) Además de su complejidad y diversidad, Big Data introduce desafíos significativos en términos de veracidad y variabilidad al requerir métodos avanzados de limpieza y validación de datos para asegurar la integridad del análisis.
5. Para la RI, el análisis de las redes puede:
 - a) Ayudar a identificar grupos de interés.
 - b) Indicar la importancia de una entidad en la transmisión de la información.
 - c) Revelar patrones de influencia dentro de una comunidad.
 - d) Ayudar a comprender la conectividad y la accesibilidad entre las entidades.
6. En el modelo de representación del conocimiento basado en herencia se puede asegurar que:
 - a) La cantidad de padres que puede tener un nodo no es mayor que 4.
 - b) Solo se usa en entornos referentes a la biología.
 - c) El razonamiento deducido está respaldado por al menos un camino dentro del grafo.
 - d) Las conclusiones pueden ser canceladas si el grafo es ambiguo.
7. Se tiene un grafo donde cada nodo es un personaje de cierto libro de cuentos y la existencia de las aristas está definida si dos personajes aparecen en el mismo cuento. Se puede asegurar que:

- a) La centralidad de grado es útil para identificar los personajes con más conexiones dentro de la red, lo que puede indicar su importancia en el libro.
 - b) La centralidad de cercanía indica el grado de conexión de cada personaje con el resto de los personajes de los cuentos del libro.
 - c) La centralidad de intermediación identifica a los personajes que actúan como conectores entre personajes de cuentos distintos.
 - d) La centralidad de cercanía ofrece una relación entre la cantidad de vecinos de un nodo con respecto a la longitud máxima de un camino dentro del grafo partiendo del nodo en cuestión.
8. El componente responsable de la gestión de recursos y planificación de tareas en Hadoop es:
- a) MapReduce.
 - b) HDFS.
 - c) Hadoop Common.
 - d) YARN.
9. Una empresa de comercio electrónico necesita mejorar su motor de búsqueda para proporcionar resultados más relevantes a sus usuarios. Actualmente, los resultados de la búsqueda no son precisos y los usuarios a menudo encuentran dificultades para encontrar productos específicos. La empresa está considerando implementar una indexación por conceptos para mejorar la relevancia de los resultados de búsqueda. ¿Qué beneficios podría provocar este cambio?
- a) Permite adaptarse fácilmente a cambios en el vocabulario y la terminología utilizada en las descripciones de los productos.
 - b) Facilita la visualización de productos al agruparlos por categorías o características comunes.
 - c) Mejora la precisión en la búsqueda de productos relacionados, incluso cuando no coinciden exactamente con los términos de búsqueda del usuario.
 - d) Ayuda a identificar automáticamente relaciones entre productos, lo que puede mejorar las recomendaciones personalizadas a los usuarios.
10. Se puede afirmar que:
- a) La cantidad de conexiones de un nodo siempre indica su influencia en la red.
 - b) Todas las relaciones en una red tienen la misma importancia para el análisis.
 - c) No es posible identificar subgrupos dentro de una red utilizando análisis de redes.
 - d) El tamaño de una red es siempre indicativo de su efectividad en la transmisión de información.
11. La Web 2.0 se caracteriza por:
- a) Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
 - b) Las plataformas ofrecen experiencias a la medida, permitiendo a los usuarios personalizar sus perfiles, recibir recomendaciones ajustadas al contenido y participar en filtrado colaborativo.
 - c) La existencia de sitios web dinámicos e interactivos que permiten a los usuarios participar, comentar e interactuar tanto con los creadores de contenido como con otros usuarios.
 - d) El uso de distintas tecnologías para crear experiencias web más interactivas y con mayor capacidad de respuesta.
12. La Web 3 se conoce como:
- a) Web de solo lectura.
 - b) Internet de las cosas.
 - c) Web de escritura-lectura.
 - d) Web semántica.
13. ¿Cuál de las siguientes opciones describe mejor la diferencia clave entre Web Crawling y Web Scraping?
- a) Web Scraping implica el análisis de la estructura y el contenido de las páginas web para extraer datos, mientras que Web Crawling se refiere a la descarga y almacenamiento de páginas web completas.
 - b) Web Scraping es más eficaz para rastrear e indexar contenido web para motores de búsqueda, mientras que Web Crawling se utiliza principalmente para la extracción de datos en proyectos de investigación.
 - c) Web Crawling se centra en la exploración y recopilación de enlaces de múltiples sitios web, mientras que Web Scraping se enfoca en la extracción específica de datos de páginas web individuales.
 - d) Web Crawling se realiza utilizando herramientas de automatización como Selenium WebDriver, mientras que Web Scraping se lleva a cabo mediante el análisis de HTML y CSS.
14. Sobre el algoritmo de PageRank, visto en clase, se puede afirmar que:
- a) Evalúa la importancia de un sitio web en función de la calidad y cantidad de enlaces entrantes que recibe de otros sitios web.
 - b) Asigna una puntuación baja a los sitios web que contienen muchos enlaces salientes, ya que indica una falta de relevancia.

- c) Asigna una puntuación alta a los sitios web que tienen un gran número de enlaces entrantes sin tener en cuenta la calidad de esos enlaces.
 - d) Solo tiene en cuenta el contenido en un sitio web para determinar su relevancia en los resultados de búsqueda.
15. El concepto de *rank sink* en el algoritmo de PageRank representa páginas web:
- a) Con una baja calidad de contenido y una cantidad insuficiente de enlaces salientes, lo que las hace menos relevantes en los resultados de búsqueda.
 - b) Con distribuciones uniformes de la puntuación de PageRank.
 - c) Con un alto número de enlaces salientes que no reciben enlaces entrantes, lo que puede afectar negativamente su puntuación de PageRank.
 - d) Que tienen una puntuación de PageRank más alta que otras debido a la manipulación de enlaces entrantes y salientes.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. La afirmación que mejor refleja el principio subyacente de PageRank, considerando su importancia en la clasificación de los sitios web, es:
 - a) El algoritmo de PageRank considera tanto la cantidad como la calidad de los enlaces entrantes, asignando mayor valor a los enlaces provenientes de sitios web considerados como “importantes”.
 - b) PageRank valora más la cantidad de enlaces entrantes a una página web, independientemente de la calidad o relevancia de estos enlaces.
 - c) La efectividad de PageRank se basa exclusivamente en el análisis de las palabras clave contenidas en los enlaces entrantes, sin considerar la estructura de enlace de la Web.
 - d) PageRank opera bajo el supuesto de que los enlaces entrantes y salientes tienen el mismo impacto en la valoración de la relevancia de una página web.
2. La Web 2.0 se caracteriza por:
 - a) Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
 - b) Las plataformas ofrecen experiencias a la medida, permitiendo a los usuarios personalizar sus perfiles, recibir recomendaciones ajustadas al contenido y participar en filtrado colaborativo.
 - c) La existencia de sitios web dinámicos e interactivos que permiten a los usuarios participar, comentar e interactuar tanto con los creadores de contenido como con otros usuarios.
 - d) El uso de distintas tecnologías para crear experiencias web más interactivas y con mayor capacidad de respuesta.
3. En la RI en el contexto de Big Data, se puede asegurar que:
 - a) El almacenamiento distribuido es una técnica obsoleta.
 - b) El almacenamiento distribuido centraliza todos los datos en un único servidor para facilitar su gestión y mantenimiento.
 - c) El almacenamiento distribuido reparte los datos en múltiples servidores para mejorar la disponibilidad y la redundancia del sistema.
 - d) El almacenamiento distribuido ofrece ventajas significativas en términos de escalabilidad y rendimiento en comparación con el almacenamiento centralizado.
4. En un SRI la indexación:
 - a) Mejora la experiencia del usuario.
 - b) Consiste en asociar un identificador único a cada dato almacenado en el sistema.
 - c) Facilita la RI relevante.
 - d) Permite la organización y la categorización de la información.
5. Para la RI, el análisis de las redes puede:
 - a) Ayudar a identificar grupos de interés.
 - b) Indicar la importancia de una entidad en la transmisión de la información.
 - c) Revelar patrones de influencia dentro de una comunidad.
 - d) Ayudar a comprender la conectividad y la accesibilidad entre las entidades.
6. ¿Qué algoritmos permiten obtener información de una red?
 - a) Índices de centralidad.
 - b) Detección de comunidades.
 - c) Hypertext Induced Topic Selection (HITS).
 - d) PageRank.
7. En una biblioteca digital se necesita implementar un sistema que permita a los usuarios encontrar libros y artículos científicos de forma rápida y precisa. Los documentos están en diversos formatos, incluyendo PDF, EPUB y HTML. Se requiere seleccionar un algoritmo de indexación adecuado para el sistema, por lo que el programador designado para la implementación debe considerar:
 - a) La capacidad para manejar documentos en diferentes formatos.
 - b) La velocidad de indexación y recuperación de datos.
 - c) La complejidad del algoritmo en términos de implementación y mantenimiento.

- d) La capacidad del algoritmo para procesar imágenes incrustadas o referenciadas en los ficheros.
8. La Web 3 se conoce como:
- Web de solo lectura.
 - Internet de las cosas.
 - Web de escritura-lectura.
 - Web semántica.
9. Dentro del análisis de redes, la centralidad de grado mide la importancia de un nodo basándose en:
- La cantidad de vecinos del nodo.
 - El número de aristas que posee el nodo.
 - La cantidad de veces que aparece el nodo en el camino mínimo entre cualquier par de nodos.
 - El grado del nodo.
10. Una buena práctica de SEO para mejorar el posicionamiento de un sitio web en los motores de búsqueda es:
- Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - Copiar contenido directamente de otros sitios web populares para aumentar la cantidad de páginas indexadas.
 - Llenar el contenido con palabras clave irrelevantes para aumentar la densidad de palabras clave.
 - Obtener enlaces de otros sitios web relevantes y de calidad que apunten al sitio.
11. Al “relajar” el concepto de clique en la detección de comunidades se intenta solucionar:
- La uniformidad de los nodos en términos de grado.
 - La necesidad de datos externos para analizar la red.
 - La imposibilidad de conectar nodos distantes.
 - El solapamiento y la complejidad computacional.
12. El algoritmo Hypertext Induced Topic Selection (HITS) intenta buscar nodos especiales. Estos son conocidos como:
- Sensibles.
 - Centrales.
 - Hubs.
 - Autoridades.
13. ¿Qué es un 'Uniform Resource Locator (URL) Frontier' en el contexto de Web Crawling?
- Una base de datos que almacena URLs únicas identificadas como recursos en la web.
 - Una técnica para filtrar URLs irrelevantes y mejorar la eficiencia del crawling.
 - Un protocolo que define cómo se deben formatear las URLs para el crawling.
 - Una lista prioritaria de URLs que aún no han sido visitadas por el crawler.
14. Dentro del ecosistema de Hadoop, el HDFS se caracteriza por:
- La capacidad ilimitada de almacenamiento.
 - La tolerancia a fallos mediante la replicación de datos.
 - El almacenamiento exclusivo para archivos de texto.
 - El modelo de acceso y de escritura de datos en tiempo real.
15. Considerando las prácticas éticas y legales en el Web Scraping, se puede asegurar que:
- Web Scraping sobre datos personales sin consentimiento es generalmente aceptado si los datos se utilizan con fines de investigación.
 - Es importante revisar y respetar los términos de servicio del sitio web, así como las leyes aplicables de protección de datos y derechos de autor, antes de realizar Web Scraping.
 - Web Scraping requiere considerar las políticas de `robots.txt` del sitio web objetivo.
 - La extracción de datos mediante Web Scraping siempre es legal, independientemente de las leyes locales sobre derechos de autor y privacidad de datos definidos en los sitios web.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Un SRI es capaz de:
 - a) No necesitar de ningún almacenamiento externo para alojar los índices de los datos.
 - b) Reducir el tiempo de indexación de los datos si utiliza servidores distribuidos dentro de la red para que cada uno ejecute la indexación del mismo conjunto de datos.
 - c) Generar índices invertidos de manera óptima sin considerar el contexto.
 - d) Crear los índices asociados a los datos sin tener que analizar cada dato.
2. En un sistema donde el conocimiento está definido a partir de reglas se puede asegurar que:
 - a) El nivel de especificidad de las reglas está limitado.
 - b) La representación de conocimiento está basada en la lógica proposicional.
 - c) El razonamiento se activa en cadena hacia delante.
 - d) El orden en que se definen las reglas no altera el razonamiento del sistema.
3. Una buena práctica de SEO para mejorar el posicionamiento de un sitio web en los motores de búsqueda es:
 - a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Copiar contenido directamente de otros sitios web populares para aumentar la cantidad de páginas indexadas.
 - c) Llenar el contenido con palabras clave irrelevantes para aumentar la densidad de palabras clave.
 - d) Obtener enlaces de otros sitios web relevantes y de calidad que apunten al sitio.
4. Analizar una red permite:
 - a) Evaluar la calidad del contenido de un sitio web.
 - b) Obtener predicciones exactas de eventos futuros en mercados financieros.
 - c) Detectar posibles tendencias antes de que se conviertan en tendencia.
 - d) Encontrar nodos “sensibles” o críticos para la red.
5. En un sistema donde el conocimiento está orientado a objetos se puede asegurar que:
 - a) Los marcos y las bandas son las estructuras utilizadas para la representación del modelo.
 - b) Las funciones de agregación dificultan poder establecer relaciones entre los objetos.
 - c) Se enfatiza la atención a la información de la cual se extrajo el conocimiento.
 - d) No es posible definir especializaciones de los objetos de la vida real dentro del sistema.
6. El componente responsable de la gestión de recursos y planificación de tareas en Hadoop es:
 - a) MapReduce.
 - b) HDFS.
 - c) Hadoop Common.
 - d) YARN.
7. Con respecto a WordNet y su contribución a las ontologías y la representación del conocimiento en los sistemas de RI, puede afirmarse que:
 - a) En WordNet, una palabra está asociada a un conjunto de sinónimos (synsets), siendo estas palabras intercambiables en un contexto.
 - b) WordNet diferencia claramente entre los significados de palabras según su uso en diferentes contextos, lo que permite aplicaciones avanzadas en desambiguación semántica más allá de los sistemas de recomendación y RI.
 - c) La integración de WordNet en sistemas de RI permite la expansión de consultas y la mejora de la precisión de los resultados al entender mejor el significado de los términos de búsqueda a través de su contexto semántico y las relaciones entre palabras.
 - d) Una limitación de WordNet en la representación del conocimiento es su enfoque en el idioma inglés, lo que plantea desafíos en la aplicación global y la interoperabilidad con sistemas de información multilingües.
8. Sobre WordNet y su aplicación en el procesamiento del lenguaje natural, se puede afirmar que:

- a) Aunque WordNet es una herramienta valiosa en el NLP, su estructura no incluye información sobre la frecuencia de uso de las palabras en el lenguaje natural.
 - b) El diseño de WordNet facilita su integración en aplicaciones multilingües de NLP, aunque su desarrollo original se centró en el inglés.
 - c) Los synsets facilitan la identificación de relaciones semánticas entre palabras, como la hiperonimia y la meronimia, enriqueciendo tareas de NLP.
 - d) WordNet proporciona una base para la desambiguación semántica al agrupar palabras con significados similares en synsets.
9. No se considera como técnica para detectar comunidades en una red:
- a) Analizar la mutualidad de los enlaces.
 - b) Usar el agrupamiento jerárquico.
 - c) Utilizar el algoritmo de K-Means.
 - d) Encontrar cliques de vértices de grado par.
10. Sobre el algoritmo de PageRank, visto en clase, se puede afirmar que:
- a) Evalúa la importancia de un sitio web en función de la calidad y cantidad de enlaces entrantes que recibe de otros sitios web.
 - b) Asigna una puntuación baja a los sitios web que contienen muchos enlaces salientes, ya que indica una falta de relevancia.
 - c) Asigna una puntuación alta a los sitios web que tienen un gran número de enlaces entrantes sin tener en cuenta la calidad de esos enlaces.
 - d) Solo tiene en cuenta el contenido en un sitio web para determinar su relevancia en los resultados de búsqueda.
11. En una red de transporte donde cada nodo es una parada de autobús y las aristas representan si existe un carro que pasa por ambos sitios, ¿qué puede mejorar el sistema de transporte?
- a) Utilizar la centralidad de intermediación para identificar las paradas de transferencia clave y establecer nuevas conexiones entre diferentes líneas de transporte público.
 - b) Utilizar el grafo inducido de los nodos con mayor valor en la centralidad de grado para aplicar la centralidad de intermediación con el propósito de reforzar las paradas con mayor tráfico.
 - c) No implementar sistemas de información en tiempo real para los usuarios, ya que pueden aumentar la carga de los trabajadores.
 - d) No considerar la centralidad de intermediación de las estaciones de transporte público al planificar rutas y horarios, ya que no tiene impacto en las conexiones entre las paradas.
12. En un sistema de control de tráfico urbano basado en reglas, ¿cuál de las siguientes reglas sería más efectiva para manejar situaciones de congestión en una intersección?
- a) Si hay pocos vehículos en la intersección, reducir el tiempo de los semáforos en verde.
 - b) Si hay muchos vehículos en la intersección, aumentar el tiempo de los semáforos en verde.
 - c) Si hay un vehículo de emergencia en la intersección, detener todos los demás vehículos.
 - d) Si hay muchos peatones cruzando la intersección, reducir el tiempo de los semáforos en rojo.
13. Una plataforma de comercio electrónico desea mejorar la experiencia del usuario al permitir una navegación más personalizada y contextualizada. Actualmente los usuarios tienen dificultades para encontrar productos relevantes debido a la gran cantidad de opciones disponibles. La empresa está interesada en implementar características de la Web 2.5 y la Web Semántica para abordar este problema. ¿Qué características podrían ayudar para ofrecer una navegación más personalizada y contextualizada?
- a) La implementación de ontologías y metadatos para enriquecer la descripción de productos y mejorar la precisión de las recomendaciones.
 - b) La optimización de la velocidad de carga del sitio web para mejorar la experiencia del usuario y reducir el abandono del carrito de compra.
 - c) La incorporación de los productos en tendencia en el mercado.
 - d) La integración de redes sociales para permitir la recomendación de productos basada en las preferencias de amigos y contactos.
14. La Web 2.0 se caracteriza por:
- a) Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
 - b) Las plataformas ofrecen experiencias a la medida, permitiendo a los usuarios personalizar sus perfiles, recibir recomendaciones ajustadas al contenido y participar en filtrado colaborativo.
 - c) La existencia de sitios web dinámicos e interactivos que permiten a los usuarios participar, comentar e interactuar tanto con los creadores de contenido como con otros usuarios.

- d) El uso de distintas tecnologías para crear experiencias web más interactivas y con mayor capacidad de respuesta.
15. La computación evolutiva:
- a) Utiliza algoritmos para buscar soluciones óptimas en grandes espacios de búsqueda, lo que la hace adecuada para problemas de optimización en la RI.
 - b) Es útil solo para problemas de clasificación de documentos y no para otras tareas de RI en general.
 - c) No es aplicable en la RI debido a la complejidad de los algoritmos evolutivos.
 - d) Solo puede manejar conjuntos de datos pequeños y no es escalable a grandes volúmenes de datos.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. El algoritmo de PageRank puede describirse como un procedimiento utilizado para:
 - a) Determinar la velocidad de carga de una página web en un navegador.
 - b) Calcular la relevancia de una página web en función de la cantidad y calidad de los enlaces que apuntan hacia ella.
 - c) Clasificar las páginas web en función de su edad y autoridad.
 - d) Prevenir el spam y el contenido no deseado en las páginas web.
2. Si una red cumple la propiedad de ser un grafo de mundo pequeño, entonces se conoce que:
 - a) La red posee pocos vértices.
 - b) La longitud media del camino entre todo par de vértices es pequeña.
 - c) El número de componentes fuertemente conexas está relacionado con la cantidad de grafos K_n presentes.
 - d) El grafo es un anillo regular de grado 5.
3. Los Web Crawlers se enfrentan a desafíos constantes. Dentro de ellos se encuentran:
 - a) La falta de acceso a la base de datos del servidor web para extraer información actualizada.
 - b) La modificación del código y la estructura del sitio web.
 - c) La incapacidad para interpretar correctamente el lenguaje de programación utilizado en el desarrollo de los sitios web.
 - d) La dificultad para generar el contenido dinámico en tiempo real.
4. Se tiene un grafo donde cada nodo es un personaje de cierto libro de cuentos y la existencia de las aristas está definida si dos personajes aparecen en el mismo cuento. Se puede asegurar que:
 - a) La centralidad de grado es útil para identificar los personajes con más conexiones dentro de la red, lo que puede indicar su importancia en el libro.
 - b) La centralidad de cercanía indica el grado de conexión de cada personaje con el resto de los personajes de los cuentos del libro.
 - c) La centralidad de intermediación identifica a los personajes que actúan como conectores entre personajes de cuentos distintos.
 - d) La centralidad de cercanía ofrece una relación entre la cantidad de vecinos de un nodo con respecto a la longitud máxima de un camino dentro del grafo partiendo del nodo en cuestión.
5. Se puede afirmar que:
 - a) Uno de los desafíos en el procesamiento de Big Data es la capacidad de gestionar y analizar datos provenientes de diversas fuentes y en diferentes formatos de manera eficiente.
 - b) MapReduce es un modelo de procesamiento distribuido utilizado para trabajar con grandes volúmenes de datos.
 - c) Los SRI pueden beneficiarse de MapReduce para mejorar la RI relevante.
 - d) El término “Big Data” se refiere exclusivamente al volumen de datos que una organización maneja, sin tener en cuenta la velocidad, la variedad y la veracidad de los datos.
6. Analizar una red permite:
 - a) Evaluar la calidad del contenido de un sitio web.
 - b) Obtener predicciones exactas de eventos futuros en mercados financieros.
 - c) Detectar posibles tendencias antes de que se conviertan en tendencia.
 - d) Encontrar nodos “sensibles” o críticos para la red.
7. La diferencia entre la indexación por tokens y la indexación por conceptos puede definirse como:
 - a) La indexación por tokens normaliza los datos reduciéndolos a su forma básica, mientras que la indexación por conceptos utiliza un método de ordenación para organizar los términos característicos de los datos.
 - b) La indexación por tokens asigna un valor numérico a cada término de los datos, mientras que la indexación por conceptos utiliza algoritmos de encriptación para proteger la privacidad de los datos.

- c) La indexación por tokens asigna pesos a los términos basados en su importancia relativa, mientras que la indexación por conceptos utiliza un sistema de etiquetado para asociar términos con características generales.
 - d) La indexación por tokens divide los datos en términos individuales, mientras que la indexación por conceptos agrupa los datos en categorías definidas.
8. En un grafo una comunidad es:
- a) Un conjunto de nodos altamente conectados que forman un subgrafo completamente independiente del resto de la red.
 - b) Un conjunto de nodos aislados.
 - c) Un conjunto de nodos que no comparten ninguna similitud estructural o funcional entre sí.
 - d) Un conjunto de nodos que están más densamente interconectados entre sí que con los nodos fuera del conjunto.
9. La centralidad de intermediación de un nodo indica:
- a) La frecuencia con la que un nodo actúa como puente en el camino más corto entre otros dos nodos.
 - b) El número total de conexiones entrantes y salientes.
 - c) La resistencia del nodo a fallos.
 - d) La cantidad de nodos vecinos directos.
10. En una biblioteca digital se necesita implementar un sistema que permita a los usuarios encontrar libros y artículos científicos de forma rápida y precisa. Los documentos están en diversos formatos, incluyendo PDF, EPUB y HTML. Se requiere seleccionar un algoritmo de indexación adecuado para el sistema, por lo que el programador designado para la implementación debe considerar:
- a) La capacidad para manejar documentos en diferentes formatos.
 - b) La velocidad de indexación y recuperación de datos.
 - c) La complejidad del algoritmo en términos de implementación y mantenimiento.
 - d) La capacidad del algoritmo para procesar imágenes incrustadas o referenciadas en los ficheros.
11. En una red de transporte donde cada nodo es una parada de autobús y las aristas representan si existe un carro que pasa por ambos sitios, ¿qué puede mejorar el sistema de transporte?
- a) Utilizar la centralidad de intermediación para identificar las paradas de transferencia clave y establecer nuevas conexiones entre diferentes líneas de transporte público.
 - b) Utilizar el grafo inducido de los nodos con mayor valor en la centralidad de grado para aplicar la centralidad de intermediación con el propósito de reforzar las paradas con mayor tráfico.
 - c) No implementar sistemas de información en tiempo real para los usuarios, ya que pueden aumentar la carga de los trabajadores.
 - d) No considerar la centralidad de intermediación de las estaciones de transporte público al planificar rutas y horarios, ya que no tiene impacto en las conexiones entre las paradas.
12. El algoritmo de PageRank converge si:
- a) Se define un factor de normalización en la fórmula de la función.
 - b) La norma de la diferencia entre los vectores es menor a un umbral predefinido.
 - c) Finaliza la ejecución cuando el número de iteraciones excede un máximo de iteraciones previamente definido.
 - d) El algoritmo no se implementa de forma iterativa.
13. En la RI en el contexto de Big Data, se puede asegurar que:
- a) El almacenamiento distribuido es una técnica obsoleta.
 - b) El almacenamiento distribuido centraliza todos los datos en un único servidor para facilitar su gestión y mantenimiento.
 - c) El almacenamiento distribuido reparte los datos en múltiples servidores para mejorar la disponibilidad y la redundancia del sistema.
 - d) El almacenamiento distribuido ofrece ventajas significativas en términos de escalabilidad y rendimiento en comparación con el almacenamiento centralizado.
14. En el contexto del procesamiento de grandes conjuntos de datos, se puede asegurar que:
- a) MapReduce no es adecuado para el procesamiento de datos no estructurados.
 - b) MapReduce solo puede manejar pequeñas cantidades de datos y no escala bien a grandes conjuntos de datos.
 - c) MapReduce es un enfoque para procesar datos de forma secuencial en un solo servidor para evitar problemas de concurrencia.

- d) MapReduce divide una tarea en múltiples pasos de map y reduce que se ejecutan de forma secuencial en diferentes servidores para mejorar el rendimiento y la escalabilidad.
15. La transición de la Web 1.0 a la Web 2.0 se caracterizó principalmente por:
- a) La reducción en el uso de HTML y CSS en el desarrollo de sitios web.
 - b) La disminución de la importancia de los motores de búsqueda en la navegación web.
 - c) El cambio de páginas web estáticas a dinámicas, permitiendo la interacción del usuario y la generación de contenido.
 - d) El aumento en la velocidad de conexión a internet, que permitió una mejor calidad de las páginas web.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Sobre WordNet y su aplicación en el procesamiento del lenguaje natural, se puede afirmar que:
 - a) Aunque WordNet es una herramienta valiosa en el NLP, su estructura no incluye información sobre la frecuencia de uso de las palabras en el lenguaje natural.
 - b) El diseño de WordNet facilita su integración en aplicaciones multilingües de NLP, aunque su desarrollo original se centró en el inglés.
 - c) Los synsets facilitan la identificación de relaciones semánticas entre palabras, como la hiperonimia y la meronimia, enriqueciendo tareas de NLP.
 - d) WordNet proporciona una base para la desambiguación semántica al agrupar palabras con significados similares en synsets.
2. La política de ordenación de URLs en los Web Crawlers tiene como aspecto fundamental:
 - a) Limitar el acceso de los crawlers a ciertas secciones de un sitio web, evitando el rastreo de URLs consideradas menos importantes o sensibles.
 - b) Establecer la prioridad de rastreo de las URLs, determinando el orden en que los crawlers visitan y procesan cada página web.
 - c) Definir la estructura de la URL de destino, asegurando que estén ordenadas alfabéticamente para facilitar la navegación y la indexación.
 - d) Determinar la forma en que los crawlers asignan un valor de relevancia a cada URL para clasificarlas en el índice de búsqueda.
3. En un SRI la indexación:
 - a) Mejora la experiencia del usuario.
 - b) Consiste en asociar un identificador único a cada dato almacenado en el sistema.
 - c) Facilita la RI relevante.
 - d) Permite la organización y la categorización de la información.
4. El concepto de *rank sink* en el algoritmo de PageRank representa páginas web:
 - a) Con una baja calidad de contenido y una cantidad insuficiente de enlaces salientes, lo que las hace menos relevantes en los resultados de búsqueda.
 - b) Con distribuciones uniformes de la puntuación de PageRank.
 - c) Con un alto número de enlaces salientes que no reciben enlaces entrantes, lo que puede afectar negativamente su puntuación de PageRank.
 - d) Que tienen una puntuación de PageRank más alta que otras debido a la manipulación de enlaces entrantes y salientes.
5. Los algoritmos para detectar comunidades en una red intentan:
 - a) Buscar subgrafos tal que no incluyan nodos cuya ausencia desconecte al subgrafo.
 - b) Seleccionar aleatoriamente nodos de alto grado y sus vecinos.
 - c) Buscar conjuntos donde cada nodo de un mismo conjunto tenga características similares al resto de los nodos del conjunto.
 - d) Encontrar grupos donde los nodos pertenecientes a los mismos grupos son cercanos bajo cierta métrica y lejanos con respecto a los nodos de otros grupos.
6. La transición de la Web 1.0 a la Web 2.0 se caracterizó principalmente por:
 - a) La reducción en el uso de HTML y CSS en el desarrollo de sitios web.
 - b) La disminución de la importancia de los motores de búsqueda en la navegación web.
 - c) El cambio de páginas web estáticas a dinámicas, permitiendo la interacción del usuario y la generación de contenido.
 - d) El aumento en la velocidad de conexión a internet, que permitió una mejor calidad de las páginas web.
7. La diferencia entre la indexación por tokens y la indexación por conceptos puede definirse como:
 - a) La indexación por tokens normaliza los datos reduciéndolos a su forma básica, mientras que la indexación por conceptos utiliza un método de ordenación para organizar los términos característicos de los datos.

- b) La indexación por tokens asigna un valor numérico a cada término de los datos, mientras que la indexación por conceptos utiliza algoritmos de encriptación para proteger la privacidad de los datos.
 - c) La indexación por tokens asigna pesos a los términos basados en su importancia relativa, mientras que la indexación por conceptos utiliza un sistema de etiquetado para asociar términos con características generales.
 - d) La indexación por tokens divide los datos en términos individuales, mientras que la indexación por conceptos agrupa los datos en categorías definidas.
8. Se puede afirmar que:
- a) Uno de los desafíos en el procesamiento de Big Data es la capacidad de gestionar y analizar datos provenientes de diversas fuentes y en diferentes formatos de manera eficiente.
 - b) MapReduce es un modelo de procesamiento distribuido utilizado para trabajar con grandes volúmenes de datos.
 - c) Los SRI pueden beneficiarse de MapReduce para mejorar la RI relevante.
 - d) El término “Big Data” se refiere exclusivamente al volumen de datos que una organización maneja, sin tener en cuenta la velocidad, la variedad y la veracidad de los datos.
9. En el contexto de la representación del conocimiento basada en herencia, ¿qué caracteriza a la herencia cancelable?
- a) Las propiedades heredadas siempre se mantienen y no pueden anularse.
 - b) La herencia es el resultado del razonamiento no transitivo.
 - c) Las conclusiones no están determinadas y dependen del nodo de interés.
 - d) No existe ambigüedad en las conclusiones obtenidas.
10. La afirmación que mejor refleja el principio subyacente de PageRank, considerando su importancia en la clasificación de los sitios web, es:
- a) El algoritmo de PageRank considera tanto la cantidad como la calidad de los enlaces entrantes, asignando mayor valor a los enlaces provenientes de sitios web considerados como “importantes”.
 - b) PageRank valora más la cantidad de enlaces entrantes a una página web, independientemente de la calidad o relevancia de estos enlaces.
 - c) La efectividad de PageRank se basa exclusivamente en el análisis de las palabras clave contenidas en los enlaces entrantes, sin considerar la estructura de enlace de la Web.
 - d) PageRank opera bajo el supuesto de que los enlaces entrantes y salientes tienen el mismo impacto en la valoración de la relevancia de una página web.
11. En el contexto del procesamiento de grandes conjuntos de datos, se puede asegurar que:
- a) MapReduce no es adecuado para el procesamiento de datos no estructurados.
 - b) MapReduce solo puede manejar pequeñas cantidades de datos y no escala bien a grandes conjuntos de datos.
 - c) MapReduce es un enfoque para procesar datos de forma secuencial en un solo servidor para evitar problemas de concurrencia.
 - d) MapReduce divide una tarea en múltiples pasos de map y reduce que se ejecutan de forma secuencial en diferentes servidores para mejorar el rendimiento y la escalabilidad.
12. Se tiene un grafo donde cada nodo es un personaje de cierto libro de cuentos y la existencia de las aristas está definida si dos personajes aparecen en el mismo cuento. Se puede asegurar que:
- a) La centralidad de grado es útil para identificar los personajes con más conexiones dentro de la red, lo que puede indicar su importancia en el libro.
 - b) La centralidad de cercanía indica el grado de conexión de cada personaje con el resto de los personajes de los cuentos del libro.
 - c) La centralidad de intermediación identifica a los personajes que actúan como conectores entre personajes de cuentos distintos.
 - d) La centralidad de cercanía ofrece una relación entre la cantidad de vecinos de un nodo con respecto a la longitud máxima de un camino dentro del grafo partiendo del nodo en cuestión.
13. Existen varias estrategias consideradas como no recomendables para mejorar el SEO de un sitio web. Dentro de estas estrategias negativas, se encuentran:
- a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Evitar el uso de las meta etiquetas y descripciones del sitio web para reflejar el contenido de la página.
 - c) Construir enlaces naturales de sitios web con autoridad y relevancia temática.

- d) Incluir una cantidad excesiva de palabras clave irrelevantes para intentar manipular los rankings de búsqueda.
14. En el contexto de la RI en redes, se puede afirmar que:
- a) La detección de comunidades ayuda a identificar grupos de nodos altamente conectados entre sí, lo que puede ser útil para comprender la estructura y el contenido de la red.
 - b) La detección de comunidades en una red siempre produce resultados objetivos y consistentes independientemente del algoritmo utilizado.
 - c) La detección de comunidades solo se aplica a redes pequeñas y simples, no a redes grandes y complejas.
 - d) La detección de comunidades no es relevante para la RI en redes, ya que se centra únicamente en la estructura de la red sin considerar el contenido.
15. Una plataforma de comercio electrónico desea mejorar la experiencia del usuario al permitir una navegación más personalizada y contextualizada. Actualmente los usuarios tienen dificultades para encontrar productos relevantes debido a la gran cantidad de opciones disponibles. La empresa está interesada en implementar características de la Web 2.5 y la Web Semántica para abordar este problema. ¿Qué características podrían ayudar para ofrecer una navegación más personalizada y contextualizada?
- a) La implementación de ontologías y metadatos para enriquecer la descripción de productos y mejorar la precisión de las recomendaciones.
 - b) La optimización de la velocidad de carga del sitio web para mejorar la experiencia del usuario y reducir el abandono del carrito de compra.
 - c) La incorporación de los productos en tendencia en el mercado.
 - d) La integración de redes sociales para permitir la recomendación de productos basada en las preferencias de amigos y contactos.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Existen varias estrategias consideradas como no recomendables para mejorar el SEO de un sitio web. Dentro de estas estrategias negativas, se encuentran:
 - a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Evitar el uso de las meta etiquetas y descripciones del sitio web para reflejar el contenido de la página.
 - c) Construir enlaces naturales de sitios web con autoridad y relevancia temática.
 - d) Incluir una cantidad excesiva de palabras clave irrelevantes para intentar manipular los rankings de búsqueda.
 2. Considerando las dimensiones y desafíos inherentes a Big Data puede afirmarse, tomando en cuenta las características clave y las implicaciones para su procesamiento y análisis, que:
 - a) Big Data se caracteriza principalmente por su pequeño volumen y uniformidad, permitiendo un procesamiento eficiente con mínimas adaptaciones de las herramientas de análisis de datos tradicionales.
 - b) Big Data no desafía la capacidad de las herramientas tradicionales de procesamiento de datos para capturar, almacenar, gestionar y analizar efectivamente la información, dada la evolución constante de las capacidades computacionales y algoritmos de optimización.
 - c) Aunque Big Data puede incluir datos estructurados, su naturaleza se expande al incorporar grandes cantidades de datos no estructurados y semiestructurados, lo que exige el uso de tecnologías especializadas en almacenamiento y procesamiento como Hadoop y sistemas de bases de datos NoSQL.
 - d) Además de su complejidad y diversidad, Big Data introduce desafíos significativos en términos de veracidad y variabilidad al requerir métodos avanzados de limpieza y validación de datos para asegurar la integridad del análisis.
 3. La transición de la Web 1.0 a la Web 2.0 se caracterizó principalmente por:
 - a) La reducción en el uso de HTML y CSS en el desarrollo de sitios web.
 - b) La disminución de la importancia de los motores de búsqueda en la navegación web.
 - c) El cambio de páginas web estáticas a dinámicas, permitiendo la interacción del usuario y la generación de contenido.
 - d) El aumento en la velocidad de conexión a internet, que permitió una mejor calidad de las páginas web.
 4. Un sistema cuenta con la siguiente información:
 - Juan pertenece a la marina.
 - Juan es capellán (sacerdote encargado de una tarea específica fuera de la parroquia).
 - Los infantes de la marina suelen ser bebedores de cerveza.
 - Un capellán no suele ser bebedor de cerveza.
 - Un bebedor de cerveza suele tener sobrepeso.
 - Por lo general, un infante de la marina está en buena condición física.
- Si se representa la información en una red de herencia se puede concluir que:
- a) Existen dos razonamientos cancelables.
 - b) Juan tiene sobrepeso.
 - c) No se puede asegurar que Juan sea capellán.
 - d) No se puede asegurar que Juan esté en buena condición física.
5. En la RI en el contexto de Big Data, se puede asegurar que:
 - a) La indexación distribuida es una técnica obsoleta.
 - b) La indexación distribuida almacena todos los datos en un solo servidor para facilitar su acceso y búsqueda.
 - c) La indexación distribuida divide los datos en múltiples fragmentos que se almacenan en varios nodos para permitir búsquedas paralelas y mejorar la escalabilidad.
 - d) La indexación distribuida no ofrece ventajas en términos de rendimiento y escalabilidad en comparación con la indexación centralizada.

6. Sobre WordNet y su aplicación en el procesamiento del lenguaje natural, se puede afirmar que:
 - a) Aunque WordNet es una herramienta valiosa en el NLP, su estructura no incluye información sobre la frecuencia de uso de las palabras en el lenguaje natural.
 - b) El diseño de WordNet facilita su integración en aplicaciones multilingües de NLP, aunque su desarrollo original se centró en el inglés.
 - c) Los synsets facilitan la identificación de relaciones semánticas entre palabras, como la hiperonimia y la meronimia, enriqueciendo tareas de NLP.
 - d) WordNet proporciona una base para la desambiguación semántica al agrupar palabras con significados similares en synsets.
7. Considerando las prácticas éticas y legales en el Web Scraping, se puede asegurar que:
 - a) Web Scraping sobre datos personales sin consentimiento es generalmente aceptado si los datos se utilizan con fines de investigación.
 - b) Es importante revisar y respetar los términos de servicio del sitio web, así como las leyes aplicables de protección de datos y derechos de autor, antes de realizar Web Scraping.
 - c) Web Scraping requiere considerar las políticas de `robots.txt` del sitio web objetivo.
 - d) La extracción de datos mediante Web Scraping siempre es legal, independientemente de las leyes locales sobre derechos de autor y privacidad de datos definidos en los sitios web.
8. En el contexto de la RI en redes, se puede afirmar que:
 - a) La detección de comunidades ayuda a identificar grupos de nodos altamente conectados entre sí, lo que puede ser útil para comprender la estructura y el contenido de la red.
 - b) La detección de comunidades en una red siempre produce resultados objetivos y consistentes independientemente del algoritmo utilizado.
 - c) La detección de comunidades solo se aplica a redes pequeñas y simples, no a redes grandes y complejas.
 - d) La detección de comunidades no es relevante para la RI en redes, ya que se centra únicamente en la estructura de la red sin considerar el contenido.
9. ¿Cuál de las siguientes opciones describe mejor la diferencia clave entre Web Crawling y Web Scraping?
 - a) Web Scraping implica el análisis de la estructura y el contenido de las páginas web para extraer datos, mientras que Web Crawling se refiere a la descarga y almacenamiento de páginas web completas.
 - b) Web Scraping es más eficaz para rastrear e indexar contenido web para motores de búsqueda, mientras que Web Crawling se utiliza principalmente para la extracción de datos en proyectos de investigación.
 - c) Web Crawling se centra en la exploración y recopilación de enlaces de múltiples sitios web, mientras que Web Scraping se enfoca en la extracción específica de datos de páginas web individuales.
 - d) Web Crawling se realiza utilizando herramientas de automatización como Selenium WebDriver, mientras que Web Scraping se lleva a cabo mediante el análisis de HTML y CSS.
10. Dentro del análisis de redes, la centralidad de grado mide la importancia de un nodo basándose en:
 - a) La cantidad de vecinos del nodo.
 - b) El número de aristas que posee el nodo.
 - c) La cantidad de veces que aparece el nodo en el camino mínimo entre cualquier par de nodos.
 - d) El grado del nodo.
11. En el algoritmo de Indexación Basada en Clasificación Bloqueada (BSBI), ¿cuál es el paso final para crear un índice invertido para la colección completa de los datos?
 - a) Dividir la colección de datos en bloques de tamaño fijo.
 - b) Eliminar los términos duplicados de los índices.
 - c) Indexar cada bloque de forma independiente.
 - d) Fusionar los índices invertidos de cada bloque.
12. No se considera como técnica para detectar comunidades en una red:
 - a) Analizar la mutualidad de los enlaces.
 - b) Usar el agrupamiento jerárquico.
 - c) Utilizar el algoritmo de K-Means.
 - d) Encontrar cliques de vértices de grado par.
13. Para transformar el contenido no estructurado de las páginas web en datos estructurados, el Web Scraping:
 - a) Convierte de forma automática imágenes a texto.
 - b) Interpreta el código JavaScript en tiempo real.
 - c) Extrae información basada en patrones de HTML/CSS.
 - d) Analiza los protocolos de red.

14. Al realizar la optimización de contenido para SEO debe considerarse:
- a) Utilizar etiquetas de título y meta descripciones únicas y relevantes para cada página.
 - b) Seleccionar las palabras clave al azar.
 - c) Crear contenido valioso y original que satisfaga las necesidades de los usuarios.
 - d) Incluir palabras clave de manera excesiva en el contenido para mejorar el posicionamiento.
15. Dentro del ecosistema de Hadoop, el HDFS se caracteriza por:
- a) La capacidad ilimitada de almacenamiento.
 - b) La tolerancia a fallos mediante la replicación de datos.
 - c) El almacenamiento exclusivo para archivos de texto.
 - d) El modelo de acceso y de escritura de datos en tiempo real.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. El componente responsable de la gestión de recursos y planificación de tareas en Hadoop es:
 - a) MapReduce.
 - b) HDFS.
 - c) Hadoop Common.
 - d) YARN.
2. La diferencia entre la indexación por tokens y la indexación por conceptos puede definirse como:
 - a) La indexación por tokens normaliza los datos reduciéndolos a su forma básica, mientras que la indexación por conceptos utiliza un método de ordenación para organizar los términos característicos de los datos.
 - b) La indexación por tokens asigna un valor numérico a cada término de los datos, mientras que la indexación por conceptos utiliza algoritmos de encriptación para proteger la privacidad de los datos.
 - c) La indexación por tokens asigna pesos a los términos basados en su importancia relativa, mientras que la indexación por conceptos utiliza un sistema de etiquetado para asociar términos con características generales.
 - d) La indexación por tokens divide los datos en términos individuales, mientras que la indexación por conceptos agrupa los datos en categorías definidas.
3. Dentro del análisis de redes, la centralidad de grado mide la importancia de un nodo basándose en:
 - a) La cantidad de vecinos del nodo.
 - b) El número de aristas que posee el nodo.
 - c) La cantidad de veces que aparece el nodo en el camino mínimo entre cualquier par de nodos.
 - d) El grado del nodo.
4. La transición de la Web 1.0 a la Web 2.0 se caracterizó principalmente por:
 - a) La reducción en el uso de HTML y CSS en el desarrollo de sitios web.
 - b) La disminución de la importancia de los motores de búsqueda en la navegación web.
 - c) El cambio de páginas web estáticas a dinámicas, permitiendo la interacción del usuario y la generación de contenido.
 - d) El aumento en la velocidad de conexión a internet, que permitió una mejor calidad de las páginas web.
5. La Web 1.0 se caracteriza por:
 - a) Los propietarios de los sitios web proporcionan contenido de forma periódica.
 - b) Uso de comunidades virtuales para popularizar los sitios web de noticias.
 - c) Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
 - d) Los sitios web se centran en brindar información en lugar de facilitar la colaboración o participación de los usuarios.
6. En una red de transporte donde cada nodo es una parada de autobús y las aristas representan si existe un carro que pasa por ambos sitios, ¿qué puede mejorar el sistema de transporte?
 - a) Utilizar la centralidad de intermediación para identificar las paradas de transferencia clave y establecer nuevas conexiones entre diferentes líneas de transporte público.
 - b) Utilizar el grafo inducido de los nodos con mayor valor en la centralidad de grado para aplicar la centralidad de intermediación con el propósito de reforzar las paradas con mayor tráfico.
 - c) No implementar sistemas de información en tiempo real para los usuarios, ya que pueden aumentar la carga de los trabajadores.
 - d) No considerar la centralidad de intermediación de las estaciones de transporte público al planificar rutas y horarios, ya que no tiene impacto en las conexiones entre las paradas.
7. ¿Qué estrategia utilizan los Web Crawlers para asegurar un rastreo eficiente y respetuoso de los recursos de los sitios web?
 - a) Extraer únicamente contenido multimedia para reducir la carga en los servidores web.
 - b) Visitar y rastrear todos los enlaces de una página web simultáneamente.

- c) Ignorar completamente el archivo `robots.txt` de los sitios web.
 - d) Seguir las directrices del archivo `robots.txt` y aplicar un retraso entre las solicitudes.
8. Un investigador necesita recopilar datos de múltiples sitios web para un estudio académico, pero se enfrenta a varios desafíos al realizar el proceso de Web Scraping de manera ética y legal. ¿Cuál de las siguientes opciones describe mejor uno de los desafíos asociados al proceso de Web Scraping?
- a) La dificultad para encontrar herramientas de Web Scraping gratuitas y fiables.
 - b) La disponibilidad limitada de datos en línea que se pueden extraer utilizando técnicas de Web Scraping.
 - c) La necesidad de estar montado sobre un crawler que cumpla con todas las políticas.
 - d) La necesidad de comprender la estructura del sitio web y su código HTML para extraer los datos correctamente.
9. En un sistema donde el conocimiento está orientado a objetos se puede asegurar que:
- a) Los marcos y las bandas son las estructuras utilizadas para la representación del modelo.
 - b) Las funciones de agregación dificultan poder establecer relaciones entre los objetos.
 - c) Se enfatiza la atención a la información de la cual se extrajo el conocimiento.
 - d) No es posible definir especializaciones de los objetos de la vida real dentro del sistema.
10. La integración de Hadoop y MapReduce en la RI trae como ventaja:
- a) La posibilidad de la extracción de información relevante y la generación de resultados significativos de grandes conjuntos de datos.
 - b) La reducción de los costos operativos a cero.
 - c) La eliminación de la necesidad de sistemas de bases de datos.
 - d) La garantía de la privacidad absoluta de los datos procesados.
11. Se tiene una red de ingredientes donde cada uno representa un nodo y las aristas simbolizan que los ingredientes forman parte de una misma receta. Se busca mejorar la experiencia culinaria mediante la elaboración de combinaciones de ingredientes más interesantes y creativas, para lo cual se debe:
- a) No tener en cuenta la centralidad de grado de los ingredientes para no darle mayor importancia a los ingredientes más comunes.
 - b) No considerar la centralidad de grado de los ingredientes, ya que todas las combinaciones de ingredientes son igualmente válidas.
 - c) Utilizar la centralidad de grado para identificar los ingredientes más populares en la red y crear combinaciones que incluyan una variedad de ingredientes menos comunes.
 - d) Utilizar la centralidad de grado para identificar los ingredientes menos conectados en la red y tomarlos en cuenta para su inclusión en las combinaciones.
12. El algoritmo de PageRank puede describirse como un procedimiento utilizado para:
- a) Determinar la velocidad de carga de una página web en un navegador.
 - b) Calcular la relevancia de una página web en función de la cantidad y calidad de los enlaces que apuntan hacia ella.
 - c) Clasificar las páginas web en función de su edad y autoridad.
 - d) Prevenir el spam y el contenido no deseado en las páginas web.
13. Considerando las dimensiones y desafíos inherentes a Big Data puede afirmarse, tomando en cuenta las características clave y las implicaciones para su procesamiento y análisis, que:
- a) Big Data se caracteriza principalmente por su pequeño volumen y uniformidad, permitiendo un procesamiento eficiente con mínimas adaptaciones de las herramientas de análisis de datos tradicionales.
 - b) Big Data no desafía la capacidad de las herramientas tradicionales de procesamiento de datos para capturar, almacenar, gestionar y analizar efectivamente la información, dada la evolución constante de las capacidades computacionales y algoritmos de optimización.
 - c) Aunque Big Data puede incluir datos estructurados, su naturaleza se expande al incorporar grandes cantidades de datos no estructurados y semiestructurados, lo que exige el uso de tecnologías especializadas en almacenamiento y procesamiento como Hadoop y sistemas de bases de datos NoSQL.
 - d) Además de su complejidad y diversidad, Big Data introduce desafíos significativos en términos de veracidad y variabilidad al requerir métodos avanzados de limpieza y validación de datos para asegurar la integridad del análisis.
14. No se considera como técnica para detectar comunidades en una red:
- a) Analizar la mutualidad de los enlaces.

- b) Usar el agrupamiento jerárquico.
 - c) Utilizar el algoritmo de K-Means.
 - d) Encontrar cliques de vértices de grado par.
15. Sobre el algoritmo de PageRank, visto en clase, se puede afirmar que:
- a) Evalúa la importancia de un sitio web en función de la calidad y cantidad de enlaces entrantes que recibe de otros sitios web.
 - b) Asigna una puntuación baja a los sitios web que contienen muchos enlaces salientes, ya que indica una falta de relevancia.
 - c) Asigna una puntuación alta a los sitios web que tienen un gran número de enlaces entrantes sin tener en cuenta la calidad de esos enlaces.
 - d) Solo tiene en cuenta el contenido en un sitio web para determinar su relevancia en los resultados de búsqueda.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Para la RI, el análisis de las redes puede:
 - a) Ayudar a identificar grupos de interés.
 - b) Indicar la importancia de una entidad en la transmisión de la información.
 - c) Revelar patrones de influencia dentro de una comunidad.
 - d) Ayudar a comprender la conectividad y la accesibilidad entre las entidades.
2. En la RI en el contexto de Big Data, se puede asegurar que:
 - a) La indexación distribuida es una técnica obsoleta.
 - b) La indexación distribuida almacena todos los datos en un solo servidor para facilitar su acceso y búsqueda.
 - c) La indexación distribuida divide los datos en múltiples fragmentos que se almacenan en varios nodos para permitir búsquedas paralelas y mejorar la escalabilidad.
 - d) La indexación distribuida no ofrece ventajas en términos de rendimiento y escalabilidad en comparación con la indexación centralizada.
3. Al realizar la optimización de contenido para SEO debe considerarse:
 - a) Utilizar etiquetas de título y meta descripciones únicas y relevantes para cada página.
 - b) Seleccionar las palabras clave al azar.
 - c) Crear contenido valioso y original que satisfaga las necesidades de los usuarios.
 - d) Incluir palabras clave de manera excesiva en el contenido para mejorar el posicionamiento.
4. El algoritmo de PageRank puede describirse como un procedimiento utilizado para:
 - a) Determinar la velocidad de carga de una página web en un navegador.
 - b) Calcular la relevancia de una página web en función de la cantidad y calidad de los enlaces que apuntan hacia ella.
 - c) Clasificar las páginas web en función de su edad y autoridad.
- d) Prevenir el spam y el contenido no deseado en las páginas web.
5. Existen varias estrategias consideradas como no recomendables para mejorar el SEO de un sitio web. Dentro de estas estrategias negativas, se encuentran:
 - a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Evitar el uso de las meta etiquetas y descripciones del sitio web para reflejar el contenido de la página.
 - c) Construir enlaces naturales de sitios web con autoridad y relevancia temática.
 - d) Incluir una cantidad excesiva de palabras clave irrelevantes para intentar manipular los rankings de búsqueda.
6. Analizar una red permite:
 - a) Evaluar la calidad del contenido de un sitio web.
 - b) Obtener predicciones exactas de eventos futuros en mercados financieros.
 - c) Detectar posibles tendencias antes de que se conviertan en tendencia.
 - d) Encontrar nodos “sensibles” o críticos para la red.
7. En un sistema de control de tráfico urbano basado en reglas, ¿cuál de las siguientes reglas sería más efectiva para manejar situaciones de congestión en una intersección?
 - a) Si hay pocos vehículos en la intersección, reducir el tiempo de los semáforos en verde.
 - b) Si hay muchos vehículos en la intersección, aumentar el tiempo de los semáforos en verde.
 - c) Si hay un vehículo de emergencia en la intersección, detener todos los demás vehículos.
 - d) Si hay muchos peatones cruzando la intersección, reducir el tiempo de los semáforos en rojo.
8. Sobre WordNet y su aplicación en el procesamiento del lenguaje natural, se puede afirmar que:

- a) Aunque WordNet es una herramienta valiosa en el NLP, su estructura no incluye información sobre la frecuencia de uso de las palabras en el lenguaje natural.
 - b) El diseño de WordNet facilita su integración en aplicaciones multilingües de NLP, aunque su desarrollo original se centró en el inglés.
 - c) Los synsets facilitan la identificación de relaciones semánticas entre palabras, como la hiperonimia y la meronimia, enriqueciendo tareas de NLP.
 - d) WordNet proporciona una base para la desambiguación semántica al agrupar palabras con significados similares en synsets.
9. La diferencia entre la indexación por tokens y la indexación por conceptos puede definirse como:
- a) La indexación por tokens normaliza los datos reduciéndolos a su forma básica, mientras que la indexación por conceptos utiliza un método de ordenación para organizar los términos característicos de los datos.
 - b) La indexación por tokens asigna un valor numérico a cada término de los datos, mientras que la indexación por conceptos utiliza algoritmos de encriptación para proteger la privacidad de los datos.
 - c) La indexación por tokens asigna pesos a los términos basados en su importancia relativa, mientras que la indexación por conceptos utiliza un sistema de etiquetado para asociar términos con características generales.
 - d) La indexación por tokens divide los datos en términos individuales, mientras que la indexación por conceptos agrupa los datos en categorías definidas.
10. La Web 2.0 se caracteriza por:
- a) Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
 - b) Las plataformas ofrecen experiencias a la medida, permitiendo a los usuarios personalizar sus perfiles, recibir recomendaciones ajustadas al contenido y participar en filtrado colaborativo.
 - c) La existencia de sitios web dinámicos e interactivos que permiten a los usuarios participar, comentar e interactuar tanto con los creadores de contenido como con otros usuarios.
 - d) El uso de distintas tecnologías para crear experiencias web más interactivas y con mayor capacidad de respuesta.
11. En el contexto del procesamiento de grandes conjuntos de datos, se puede asegurar que:
- a) MapReduce no es adecuado para el procesamiento de datos no estructurados.
 - b) MapReduce solo puede manejar pequeñas cantidades de datos y no escala bien a grandes conjuntos de datos.
 - c) MapReduce es un enfoque para procesar datos de forma secuencial en un solo servidor para evitar problemas de concurrencia.
 - d) MapReduce divide una tarea en múltiples pasos de map y reduce que se ejecutan de forma secuencial en diferentes servidores para mejorar el rendimiento y la escalabilidad.
12. En una biblioteca digital se necesita implementar un sistema que permita a los usuarios encontrar libros y artículos científicos de forma rápida y precisa. Los documentos están en diversos formatos, incluyendo PDF, EPUB y HTML. Se requiere seleccionar un algoritmo de indexación adecuado para el sistema, por lo que el programador designado para la implementación debe considerar:
- a) La capacidad para manejar documentos en diferentes formatos.
 - b) La velocidad de indexación y recuperación de datos.
 - c) La complejidad del algoritmo en términos de implementación y mantenimiento.
 - d) La capacidad del algoritmo para procesar imágenes incrustadas o referenciadas en los ficheros.
13. En el diseño de un videojuego de roles (RPG) en el que los personajes tienen atributos como salud, fuerza y velocidad, ¿cuál de las siguientes opciones representa mejor una implementación de la representación del conocimiento orientado a objetos?
- a) Cada personaje se representa como un objeto con propiedades como salud, fuerza y velocidad, y métodos para modificar y consultar estos valores.
 - b) Cada personaje se representa como una función que calcula sus atributos en función de su nivel y experiencia.
 - c) Cada personaje se representa como una matriz de números que almacena sus valores de atributos.
 - d) Cada personaje se representa como una lista de cadenas de texto que describen sus características físicas y habilidades.
14. El algoritmo Hypertext Induced Topic Selection (HITS) intenta buscar nodos especiales. Estos son conocidos como:

- a) Sensibles.
 - b) Centrales.
 - c) Hubs.
 - d) Autoridades.
15. Considerando las prácticas éticas y legales en el Web Scraping, se puede asegurar que:
- a) Web Scraping sobre datos personales sin consentimiento es generalmente aceptado si los datos se utilizan con fines de investigación.
 - b) Es importante revisar y respetar los términos de servicio del sitio web, así como las leyes aplicables de protección de datos y derechos de autor, antes de realizar Web Scraping.
 - c) Web Scraping requiere considerar las políticas de `robots.txt` del sitio web objetivo.
 - d) La extracción de datos mediante Web Scraping siempre es legal, independientemente de las leyes locales sobre derechos de autor y privacidad de datos definidos en los sitios web.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. El componente responsable de la gestión de recursos y planificación de tareas en Hadoop es:
 - a) MapReduce.
 - b) HDFS.
 - c) Hadoop Common.
 - d) YARN.
2. Un investigador necesita recopilar datos de múltiples sitios web para un estudio académico, pero se enfrenta a varios desafíos al realizar el proceso de Web Scraping de manera ética y legal. ¿Cuál de las siguientes opciones describe mejor uno de los desafíos asociados al proceso de Web Scraping?
 - a) La dificultad para encontrar herramientas de Web Scraping gratuitas y fiables.
 - b) La disponibilidad limitada de datos en línea que se pueden extraer utilizando técnicas de Web Scraping.
 - c) La necesidad de estar montado sobre un crawler que cumpla con todas las políticas.
 - d) La necesidad de comprender la estructura del sitio web y su código HTML para extraer los datos correctamente.
3. En el contexto de la representación del conocimiento basada en herencia, ¿qué caracteriza a la herencia cancelable?
 - a) Las propiedades heredadas siempre se mantienen y no pueden anularse.
 - b) La herencia es el resultado del razonamiento no transitivo.
 - c) Las conclusiones no están determinadas y dependen del nodo de interés.
 - d) No existe ambigüedad en las conclusiones obtenidas.
4. Existen varias estrategias consideradas como no recomendables para mejorar el SEO de un sitio web. Dentro de estas estrategias negativas, se encuentran:
 - a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Evitar el uso de las meta etiquetas y descripciones del sitio web para reflejar el contenido de la página.
 - c) Construir enlaces naturales de sitios web con autoridad y relevancia temática.
 - d) Incluir una cantidad excesiva de palabras clave irrelevantes para intentar manipular los rankings de búsqueda.
5. El conjunto frontera de URLs en un Web Crawler:
 - a) Delimita los sitios web que visitará en futuras iteraciones del proceso.
 - b) Indica las secciones que no pueden ser visitadas de cada sitio web.
 - c) Almacena hipervínculos que pertenecen al mismo dominio del conjunto semilla de URLs con que inició el crawler.
 - d) Es similar a un conjunto de URLs que esperan ser visitadas.
6. Los algoritmos para detectar comunidades en una red intentan:
 - a) Buscar subgrafos tal que no incluyan nodos cuya ausencia desconecte al subgrafo.
 - b) Seleccionar aleatoriamente nodos de alto grado y sus vecinos.
 - c) Buscar conjuntos donde cada nodo de un mismo conjunto tenga características similares al resto de los nodos del conjunto.
 - d) Encontrar grupos donde los nodos pertenecientes a los mismos grupos son cercanos bajo cierta métrica y lejanos con respecto a los nodos de otros grupos.
7. En el contexto de la RI en redes sociales se puede afirmar que:
 - a) El análisis de redes sociales se centra exclusivamente en la cantidad de seguidores que tiene un usuario en particular para determinar su influencia en la red.
 - b) El análisis de redes sociales no es útil para comprender la difusión de información en una red social específica, para ello se utiliza la medida de centralidad de vector propio.

- c) El análisis de centralidad de intermediación se utiliza para identificar usuarios que son importantes en una red social debido a su posición como “puentes” entre diferentes grupos de usuarios.
- d) El análisis de sentimientos se utiliza para determinar la popularidad de una publicación en redes sociales sin tener en cuenta la opinión de los usuarios.
8. Si se tiene el conjunto de páginas interconectadas $\{A \rightarrow B, C; B \rightarrow C; C \rightarrow A\}$, entonces la página con valor más alto de PageRank es:
- B.
 - C.
 - A y B.
 - A.
9. En el diseño de un videojuego de roles (RPG) en el que los personajes tienen atributos como salud, fuerza y velocidad, ¿cuál de las siguientes opciones representa mejor una implementación de la representación del conocimiento orientado a objetos?
- Cada personaje se representa como un objeto con propiedades como salud, fuerza y velocidad, y métodos para modificar y consultar estos valores.
 - Cada personaje se representa como una función que calcula sus atributos en función de su nivel y experiencia.
 - Cada personaje se representa como una matriz de números que almacena sus valores de atributos.
 - Cada personaje se representa como una lista de cadenas de texto que describen sus características físicas y habilidades.
10. En un grafo una comunidad es:
- Un conjunto de nodos altamente conectados que forman un subgrafo completamente independiente del resto de la red.
 - Un conjunto de nodos aislados.
 - Un conjunto de nodos que no comparten ninguna similitud estructural o funcional entre sí.
 - Un conjunto de nodos que están más densamente interconectados entre sí que con los nodos fuera del conjunto.
11. La Web 3 se conoce como:
- Web de solo lectura.
 - Internet de las cosas.
 - Web de escritura-lectura.
 - Web semántica.
12. La centralidad de intermediación de un nodo indica:
- La frecuencia con la que un nodo actúa como puente en el camino más corto entre otros dos nodos.
 - El número total de conexiones entrantes y salientes.
 - La resistencia del nodo a fallos.
 - La cantidad de nodos vecinos directos.
13. Una empresa de comercio electrónico necesita mejorar su motor de búsqueda para proporcionar resultados más relevantes a sus usuarios. Actualmente, los resultados de la búsqueda no son precisos y los usuarios a menudo encuentran dificultades para encontrar productos específicos. La empresa está considerando implementar una indexación por conceptos para mejorar la relevancia de los resultados de búsqueda. ¿Qué beneficios podría provocar este cambio?
- Permite adaptarse fácilmente a cambios en el vocabulario y la terminología utilizada en las descripciones de los productos.
 - Facilita la visualización de productos al agruparlos por categorías o características comunes.
 - Mejora la precisión en la búsqueda de productos relacionados, incluso cuando no coinciden exactamente con los términos de búsqueda del usuario.
 - Ayuda a identificar automáticamente relaciones entre productos, lo que puede mejorar las recomendaciones personalizadas a los usuarios.
14. Dentro del ecosistema de Hadoop, el HDFS se caracteriza por:
- La capacidad ilimitada de almacenamiento.
 - La tolerancia a fallos mediante la replicación de datos.
 - El almacenamiento exclusivo para archivos de texto.
 - El modelo de acceso y de escritura de datos en tiempo real.
15. Para la RI, el análisis de las redes puede:
- Ayudar a identificar grupos de interés.
 - Indicar la importancia de una entidad en la transmisión de la información.
 - Revelar patrones de influencia dentro de una comunidad.
 - Ayudar a comprender la conectividad y la accesibilidad entre las entidades.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. La integración de Hadoop y MapReduce en la RI trae como ventaja:
 - a) La posibilidad de la extracción de información relevante y la generación de resultados significativos de grandes conjuntos de datos.
 - b) La reducción de los costos operativos a cero.
 - c) La eliminación de la necesidad de sistemas de bases de datos.
 - d) La garantía de la privacidad absoluta de los datos procesados.
2. La centralidad de intermediación de un nodo indica:
 - a) La frecuencia con la que un nodo actúa como puente en el camino más corto entre otros dos nodos.
 - b) El número total de conexiones entrantes y salientes.
 - c) La resistencia del nodo a fallos.
 - d) La cantidad de nodos vecinos directos.
3. En el modelo de representación del conocimiento basado en herencia se puede asegurar que:
 - a) La cantidad de padres que puede tener un nodo no es mayor que 4.
 - b) Solo se usa en entornos referentes a la biología.
 - c) El razonamiento deducido está respaldado por al menos un camino dentro del grafo.
 - d) Las conclusiones pueden ser canceladas si el grafo es ambiguo.
4. Los Web Crawlers se enfrentan a desafíos constantes. Dentro de ellos se encuentran:
 - a) La falta de acceso a la base de datos del servidor web para extraer información actualizada.
 - b) La modificación del código y la estructura del sitio web.
 - c) La incapacidad para interpretar correctamente el lenguaje de programación utilizado en el desarrollo de los sitios web.
 - d) La dificultad para generar el contenido dinámico en tiempo real.
5. La computación evolutiva:
 - a) Utiliza algoritmos para buscar soluciones óptimas en grandes espacios de búsqueda, lo que la hace adecuada para problemas de optimización en la RI.
 - b) Es útil solo para problemas de clasificación de documentos y no para otras tareas de RI en general.
 - c) No es aplicable en la RI debido a la complejidad de los algoritmos evolutivos.
 - d) Solo puede manejar conjuntos de datos pequeños y no es escalable a grandes volúmenes de datos.
6. Una plataforma de comercio electrónico desea mejorar la experiencia del usuario al permitir una navegación más personalizada y contextualizada. Actualmente los usuarios tienen dificultades para encontrar productos relevantes debido a la gran cantidad de opciones disponibles. La empresa está interesada en implementar características de la Web 2.5 y la Web Semántica para abordar este problema. ¿Qué características podrían ayudar para ofrecer una navegación más personalizada y contextualizada?
 - a) La implementación de ontologías y metadatos para enriquecer la descripción de productos y mejorar la precisión de las recomendaciones.
 - b) La optimización de la velocidad de carga del sitio web para mejorar la experiencia del usuario y reducir el abandono del carrito de compra.
 - c) La incorporación de los productos en tendencia en el mercado.
 - d) La integración de redes sociales para permitir la recomendación de productos basada en las preferencias de amigos y contactos.
7. Referente al Web Crawler puede afirmarse que:
 - a) Los hipervínculos encontrados en cada sitio web que no pertenecen al dominio donde fueron encontrados se desechan, puesto que no expande el conjunto de URLs sin visitar.
 - b) Las páginas visitadas no se procesan nunca más.
 - c) No necesita de un conjunto inicial de URLs para recorrer la Web.
 - d) No tiene como objetivo indexar y recopilar información de diferentes sitios web.

8. Sobre el algoritmo de PageRank, visto en clase, se puede afirmar que:
 - a) Evalúa la importancia de un sitio web en función de la calidad y cantidad de enlaces entrantes que recibe de otros sitios web.
 - b) Asigna una puntuación baja a los sitios web que contienen muchos enlaces salientes, ya que indica una falta de relevancia.
 - c) Asigna una puntuación alta a los sitios web que tienen un gran número de enlaces entrantes sin tener en cuenta la calidad de esos enlaces.
 - d) Solo tiene en cuenta el contenido en un sitio web para determinar su relevancia en los resultados de búsqueda.
9. En un SRI la indexación:
 - a) Mejora la experiencia del usuario.
 - b) Consiste en asociar un identificador único a cada dato almacenado en el sistema.
 - c) Facilita la RI relevante.
 - d) Permite la organización y la categorización de la información.
10. Se tiene una red de ingredientes donde cada uno representa un nodo y las aristas simbolizan que los ingredientes forman parte de una misma receta. Se busca mejorar la experiencia culinaria mediante la elaboración de combinaciones de ingredientes más interesantes y creativas, para lo cual se debe:
 - a) No tener en cuenta la centralidad de grado de los ingredientes para no darle mayor importancia a los ingredientes más comunes.
 - b) No considerar la centralidad de grado de los ingredientes, ya que todas las combinaciones de ingredientes son igualmente válidas.
 - c) Utilizar la centralidad de grado para identificar los ingredientes más populares en la red y crear combinaciones que incluyan una variedad de ingredientes menos comunes.
 - d) Utilizar la centralidad de grado para identificar los ingredientes menos conectados en la red y tomarlos en cuenta para su inclusión en las combinaciones.
11. Al “relajar” el concepto de clique en la detección de comunidades se intenta solucionar:
 - a) La uniformidad de los nodos en términos de grado.
 - b) La necesidad de datos externos para analizar la red.
 - c) La imposibilidad de conectar nodos distantes.
 - d) El solapamiento y la complejidad computacional.
12. El concepto de *rank sink* en el algoritmo de PageRank representa páginas web:
 - a) Con una baja calidad de contenido y una cantidad insuficiente de enlaces salientes, lo que las hace menos relevantes en los resultados de búsqueda.
 - b) Con distribuciones uniformes de la puntuación de PageRank.
 - c) Con un alto número de enlaces salientes que no reciben enlaces entrantes, lo que puede afectar negativamente su puntuación de PageRank.
 - d) Que tienen una puntuación de PageRank más alta que otras debido a la manipulación de enlaces entrantes y salientes.
13. Existen varias estrategias consideradas como no recomendables para mejorar el SEO de un sitio web. Dentro de estas estrategias negativas, se encuentran:
 - a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Evitar el uso de las meta etiquetas y descripciones del sitio web para reflejar el contenido de la página.
 - c) Construir enlaces naturales de sitios web con autoridad y relevancia temática.
 - d) Incluir una cantidad excesiva de palabras clave irrelevantes para intentar manipular los rankings de búsqueda.
14. En el contexto de la RI en redes sociales se puede afirmar que:
 - a) El análisis de redes sociales se centra exclusivamente en la cantidad de seguidores que tiene un usuario en particular para determinar su influencia en la red.
 - b) El análisis de redes sociales no es útil para comprender la difusión de información en una red social específica, para ello se utiliza la medida de centralidad de vector propio.
 - c) El análisis de centralidad de intermediación se utiliza para identificar usuarios que son importantes en una red social debido a su posición como “puentes” entre diferentes grupos de usuarios.
 - d) El análisis de sentimientos se utiliza para determinar la popularidad de una publicación en redes sociales sin tener en cuenta la opinión de los usuarios.
15. La afirmación que mejor describe la política de amabilidad en los Web Crawlers es:

- a) Los Web Crawlers son libres de recopilar datos de cualquier sitio web sin restricciones, independientemente de la cantidad de tráfico que generen.
- b) La política de amabilidad establece pautas y reglas sobre cómo los crawlers deben interactuar con los sitios web para minimizar la carga del servidor y respetar las directivas de los administradores del sitio.
- c) La política de amabilidad de los Web Crawlers dicta que los crawlers deben priorizar ciertos tipos de contenido sobre otros, ignorando completamente ciertas páginas web.
- d) Los crawlers se diseñan para acceder a sitios web sin restricciones y extraer datos de manera agresiva para su indexación.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Para contribuir positivamente al posicionamiento orgánico de un sitio web en los motores de búsqueda se puede:
 - a) Crear contenido relevante y de alta calidad que satisfaga las necesidades de información de los usuarios.
 - b) Obtener enlaces entrantes de otros sitios web de alta autoridad y relevancia temática.
 - c) Incluir una densidad alta de palabras clave para asegurar que el sitio web aparezca en tantas búsquedas como sea posible.
 - d) Mejorar la velocidad de carga del sitio web y asegurar que sea *responsive* y fácil de usar en dispositivos móviles.
2. En el contexto de la representación del conocimiento basada en herencia, ¿qué caracteriza a la herencia cancelable?
 - a) Las propiedades heredadas siempre se mantienen y no pueden anularse.
 - b) La herencia es el resultado del razonamiento no transitivo.
 - c) Las conclusiones no están determinadas y dependen del nodo de interés.
 - d) No existe ambigüedad en las conclusiones obtenidas.
3. En un grafo una comunidad es:
 - a) Un conjunto de nodos altamente conectados que forman un subgrafo completamente independiente del resto de la red.
 - b) Un conjunto de nodos aislados.
 - c) Un conjunto de nodos que no comparten ninguna similitud estructural o funcional entre sí.
 - d) Un conjunto de nodos que están más densamente interconectados entre sí que con los nodos fuera del conjunto.
4. La afirmación que mejor refleja el principio subyacente de PageRank, considerando su importancia en la clasificación de los sitios web, es:
 - a) El algoritmo de PageRank considera tanto la cantidad como la calidad de los enlaces entrantes, asignando mayor valor a los enlaces provenientes de sitios web considerados como “importantes”.
5. Considerando las dimensiones y desafíos inherentes a Big Data puede afirmarse, tomando en cuenta las características clave y las implicaciones para su procesamiento y análisis, que:
 - a) Big Data se caracteriza principalmente por su pequeño volumen y uniformidad, permitiendo un procesamiento eficiente con mínimas adaptaciones de las herramientas de análisis de datos tradicionales.
 - b) Big Data no desafía la capacidad de las herramientas tradicionales de procesamiento de datos para capturar, almacenar, gestionar y analizar efectivamente la información, dada la evolución constante de las capacidades computacionales y algoritmos de optimización.
 - c) Aunque Big Data puede incluir datos estructurados, su naturaleza se expande al incorporar grandes cantidades de datos no estructurados y semiestructurados, lo que exige el uso de tecnologías especializadas en almacenamiento y procesamiento como Hadoop y sistemas de bases de datos NoSQL.
 - d) Además de su complejidad y diversidad, Big Data introduce desafíos significativos en términos de veracidad y variabilidad al requerir métodos avanzados de limpieza y validación de datos para asegurar la integridad del análisis.
6. La centralidad de intermediación de un nodo indica:
 - a) La frecuencia con la que un nodo actúa como puente en el camino más corto entre otros dos nodos.
5. b) PageRank valora más la cantidad de enlaces entrantes a una página web, independientemente de la calidad o relevancia de estos enlaces.
5. c) La efectividad de PageRank se basa exclusivamente en el análisis de las palabras clave contenidas en los enlaces entrantes, sin considerar la estructura de enlace de la Web.
5. d) PageRank opera bajo el supuesto de que los enlaces entrantes y salientes tienen el mismo impacto en la valoración de la relevancia de una página web.

- b) El número total de conexiones entrantes y salientes.
 - c) La resistencia del nodo a fallos.
 - d) La cantidad de nodos vecinos directos.
7. En el algoritmo de Indexación Basada en Clasificación Bloqueada (BSBI), ¿cuál es el paso final para crear un índice invertido para la colección completa de los datos?
- a) Dividir la colección de datos en bloques de tamaño fijo.
 - b) Eliminar los términos duplicados de los índices.
 - c) Indexar cada bloque de forma independiente.
 - d) Fusionar los índices invertidos de cada bloque.
8. La diferencia entre la indexación por tokens y la indexación por conceptos puede definirse como:
- a) La indexación por tokens normaliza los datos reduciéndolos a su forma básica, mientras que la indexación por conceptos utiliza un método de ordenación para organizar los términos característicos de los datos.
 - b) La indexación por tokens asigna un valor numérico a cada término de los datos, mientras que la indexación por conceptos utiliza algoritmos de encriptación para proteger la privacidad de los datos.
 - c) La indexación por tokens asigna pesos a los términos basados en su importancia relativa, mientras que la indexación por conceptos utiliza un sistema de etiquetado para asociar términos con características generales.
 - d) La indexación por tokens divide los datos en términos individuales, mientras que la indexación por conceptos agrupa los datos en categorías definidas.
9. Dentro del análisis de redes, la centralidad de grado mide la importancia de un nodo basándose en:
- a) La cantidad de vecinos del nodo.
 - b) El número de aristas que posee el nodo.
 - c) La cantidad de veces que aparece el nodo en el camino mínimo entre cualquier par de nodos.
 - d) El grado del nodo.
10. ¿Qué algoritmos permiten obtener información de una red?
- a) Índices de centralidad.
 - b) Detección de comunidades.
 - c) Hypertext Induced Topic Selection (HITS).
 - d) PageRank.
11. Dentro del ecosistema de Hadoop, el HDFS se caracteriza por:
- a) La capacidad ilimitada de almacenamiento.
 - b) La tolerancia a fallos mediante la replicación de datos.
 - c) El almacenamiento exclusivo para archivos de texto.
 - d) El modelo de acceso y de escritura de datos en tiempo real.
12. La política de ordenación de URLs en los Web Crawlers tiene como aspecto fundamental:
- a) Limitar el acceso de los crawlers a ciertas secciones de un sitio web, evitando el rastreo de URLs consideradas menos importantes o sensibles.
 - b) Establecer la prioridad de rastreo de las URLs, determinando el orden en que los crawlers visitan y procesan cada página web.
 - c) Definir la estructura de la URL de destino, asegurando que estén ordenadas alfabéticamente para facilitar la navegación y la indexación.
 - d) Determinar la forma en que los crawlers asignan un valor de relevancia a cada URL para clasificarlas en el índice de búsqueda.
13. La transición de la Web 1.0 a la Web 2.0 se caracterizó principalmente por:
- a) La reducción en el uso de HTML y CSS en el desarrollo de sitios web.
 - b) La disminución de la importancia de los motores de búsqueda en la navegación web.
 - c) El cambio de páginas web estáticas a dinámicas, permitiendo la interacción del usuario y la generación de contenido.
 - d) El aumento en la velocidad de conexión a internet, que permitió una mejor calidad de las páginas web.
14. Un SRI es capaz de:
- a) No necesitar de ningún almacenamiento externo para alojar los índices de los datos.
 - b) Reducir el tiempo de indexación de los datos si utiliza servidores distribuidos dentro de la red para que cada uno ejecute la indexación del mismo conjunto de datos.
 - c) Generar índices invertidos de manera óptima sin considerar el contexto.
 - d) Crear los índices asociados a los datos sin tener que analizar cada dato.
15. En un sistema de control de tráfico urbano basado en reglas, ¿cuál de las siguientes reglas sería más efectiva para manejar situaciones de congestión en una intersección?
- a) Si hay pocos vehículos en la intersección, reducir el tiempo de los semáforos en verde.

- b) Si hay muchos vehículos en la intersección, aumentar el tiempo de los semáforos en verde.
- c) Si hay un vehículo de emergencia en la intersección, detener todos los demás vehículos.
- d) Si hay muchos peatones cruzando la intersección, reducir el tiempo de los semáforos en rojo.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Si se tiene el conjunto de páginas interconectadas $\{A \rightarrow B, C; B \rightarrow C; C \rightarrow A\}$, entonces la página con valor más alto de PageRank es:
 - a) B.
 - b) C.
 - c) A y B.
 - d) A.
2. En la RI en el contexto de Big Data, se puede asegurar que:
 - a) El almacenamiento distribuido es una técnica obsoleta.
 - b) El almacenamiento distribuido centraliza todos los datos en un único servidor para facilitar su gestión y mantenimiento.
 - c) El almacenamiento distribuido reparte los datos en múltiples servidores para mejorar la disponibilidad y la redundancia del sistema.
 - d) El almacenamiento distribuido ofrece ventajas significativas en términos de escalabilidad y rendimiento en comparación con el almacenamiento centralizado.
3. En el algoritmo de Indexación Basada en Clasificación Bloqueada (BSBI), ¿cuál es el paso final para crear un índice invertido para la colección completa de los datos?
 - a) Dividir la colección de datos en bloques de tamaño fijo.
 - b) Eliminar los términos duplicados de los índices.
 - c) Indexar cada bloque de forma independiente.
 - d) Fusionar los índices invertidos de cada bloque.
4. Si una red cumple la propiedad de ser un grafo de mundo pequeño, entonces se conoce que:
 - a) La red posee pocos vértices.
 - b) La longitud media del camino entre todo par de vértices es pequeña.
 - c) El número de componentes fuertemente conexas está relacionado con la cantidad de grafos K_n presentes.
 - d) El grafo es un anillo regular de grado 5.
5. Al realizar la optimización de contenido para SEO debe considerarse:
 - a) Utilizar etiquetas de título y meta descripciones únicas y relevantes para cada página.
 - b) Seleccionar las palabras clave al azar.
 - c) Crear contenido valioso y original que satisfaga las necesidades de los usuarios.
 - d) Incluir palabras clave de manera excesiva en el contenido para mejorar el posicionamiento.
6. ¿Qué es un 'Uniform Resource Locator (URL) Frontier' en el contexto de Web Crawling?
 - a) Una base de datos que almacena URLs únicas identificadas como recursos en la web.
 - b) Una técnica para filtrar URLs irrelevantes y mejorar la eficiencia del crawling.
 - c) Un protocolo que define cómo se deben formatear las URLs para el crawling.
 - d) Una lista prioritaria de URLs que aún no han sido visitadas por el crawler.
7. Una plataforma de comercio electrónico desea mejorar la experiencia del usuario al permitir una navegación más personalizada y contextualizada. Actualmente los usuarios tienen dificultades para encontrar productos relevantes debido a la gran cantidad de opciones disponibles. La empresa está interesada en implementar características de la Web 2.5 y la Web Semántica para abordar este problema. ¿Qué características podrían ayudar para ofrecer una navegación más personalizada y contextualizada?
 - a) La implementación de ontologías y metadatos para enriquecer la descripción de productos y mejorar la precisión de las recomendaciones.
 - b) La optimización de la velocidad de carga del sitio web para mejorar la experiencia del usuario y reducir el abandono del carrito de compra.
 - c) La incorporación de los productos en tendencia en el mercado.
 - d) La integración de redes sociales para permitir la recomendación de productos basada en las preferencias de amigos y contactos.
8. Se puede afirmar que:

- a) Uno de los desafíos en el procesamiento de Big Data es la capacidad de gestionar y analizar datos provenientes de diversas fuentes y en diferentes formatos de manera eficiente.
 - b) MapReduce es un modelo de procesamiento distribuido utilizado para trabajar con grandes volúmenes de datos.
 - c) Los SRI pueden beneficiarse de MapReduce para mejorar la RI relevante.
 - d) El término “Big Data” se refiere exclusivamente al volumen de datos que una organización maneja, sin tener en cuenta la velocidad, la variedad y la veracidad de los datos.
9. El posicionamiento de un sitio web en los motores de búsqueda puede ser afectado por:
- a) Mantener una estructura de URL clara y coherente.
 - b) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - c) Contenido sin valor en el sitio web.
 - d) Obtener enlaces de sitios web irrelevantes y de baja calidad.
10. El concepto de *rank sink* en el algoritmo de PageRank representa páginas web:
- a) Con una baja calidad de contenido y una cantidad insuficiente de enlaces salientes, lo que las hace menos relevantes en los resultados de búsqueda.
 - b) Con distribuciones uniformes de la puntuación de PageRank.
 - c) Con un alto número de enlaces salientes que no reciben enlaces entrantes, lo que puede afectar negativamente su puntuación de PageRank.
 - d) Que tienen una puntuación de PageRank más alta que otras debido a la manipulación de enlaces entrantes y salientes.
11. Al “relajar” el concepto de clique en la detección de comunidades se intenta solucionar:
- a) La uniformidad de los nodos en términos de grado.
 - b) La necesidad de datos externos para analizar la red.
 - c) La imposibilidad de conectar nodos distantes.
 - d) El solapamiento y la complejidad computacional.
12. Se tiene un grafo donde cada nodo es un personaje de cierto libro de cuentos y la existencia de las aristas está definida si dos personajes aparecen en el mismo cuento. Se puede asegurar que:
- a) La centralidad de grado es útil para identificar los personajes con más conexiones dentro de la red, lo que puede indicar su importancia en el libro.
 - b) La centralidad de cercanía indica el grado de conexión de cada personaje con el resto de los personajes de los cuentos del libro.
 - c) La centralidad de intermediación identifica a los personajes que actúan como conectores entre personajes de cuentos distintos.
 - d) La centralidad de cercanía ofrece una relación entre la cantidad de vecinos de un nodo con respecto a la longitud máxima de un camino dentro del grafo partiendo del nodo en cuestión.
13. La afirmación que mejor refleja el principio subyacente de PageRank, considerando su importancia en la clasificación de los sitios web, es:
- a) El algoritmo de PageRank considera tanto la cantidad como la calidad de los enlaces entrantes, asignando mayor valor a los enlaces provenientes de sitios web considerados como “importantes”.
 - b) PageRank valora más la cantidad de enlaces entrantes a una página web, independientemente de la calidad o relevancia de estos enlaces.
 - c) La efectividad de PageRank se basa exclusivamente en el análisis de las palabras clave contenidas en los enlaces entrantes, sin considerar la estructura de enlace de la Web.
 - d) PageRank opera bajo el supuesto de que los enlaces entrantes y salientes tienen el mismo impacto en la valoración de la relevancia de una página web.
14. Dentro del análisis de redes, la centralidad de grado mide la importancia de un nodo basándose en:
- a) La cantidad de vecinos del nodo.
 - b) El número de aristas que posee el nodo.
 - c) La cantidad de veces que aparece el nodo en el camino mínimo entre cualquier par de nodos.
 - d) El grado del nodo.
15. En el modelo de representación del conocimiento basado en herencia se puede asegurar que:
- a) La cantidad de padres que puede tener un nodo no es mayor que 4.
 - b) Solo se usa en entornos referentes a la biología.
 - c) El razonamiento deducido está respaldado por al menos un camino dentro del grafo.
 - d) Las conclusiones pueden ser canceladas si el grafo es ambiguo.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Un SRI es capaz de:
 - a) No necesitar de ningún almacenamiento externo para alojar los índices de los datos.
 - b) Reducir el tiempo de indexación de los datos si utiliza servidores distribuidos dentro de la red para que cada uno ejecute la indexación del mismo conjunto de datos.
 - c) Generar índices invertidos de manera óptima sin considerar el contexto.
 - d) Crear los índices asociados a los datos sin tener que analizar cada dato.
2. Dentro del ecosistema de Hadoop, el HDFS se caracteriza por:
 - a) La capacidad ilimitada de almacenamiento.
 - b) La tolerancia a fallos mediante la replicación de datos.
 - c) El almacenamiento exclusivo para archivos de texto.
 - d) El modelo de acceso y de escritura de datos en tiempo real.
3. La Web 3 se conoce como:
 - a) Web de solo lectura.
 - b) Internet de las cosas.
 - c) Web de escritura-lectura.
 - d) Web semántica.
4. El propósito de la política de revisitado en los Web Crawlers es:
 - a) Determinar la frecuencia con la que los crawlers deben volver a visitar una URL específica para mantener la información actualizada en el índice de búsqueda.
 - b) Limitar el acceso de los crawlers a ciertos servidores luego de visitar las páginas alojadas en estos.
 - c) Establecer reglas sobre el tiempo máximo que los crawlers pueden pasar en un sitio web durante cada visita para evitar sobrecargar los servidores.
 - d) Definir la prioridad de rastreo de las URLs, determinando el orden en que los crawlers visitan y procesan cada página web.
5. Considerando las dimensiones y desafíos inherentes a Big Data puede afirmarse, tomando en cuenta las características clave y las implicaciones para su procesamiento y análisis, que:
 - a) Big Data se caracteriza principalmente por su pequeño volumen y uniformidad, permitiendo un procesamiento eficiente con mínimas adaptaciones de las herramientas de análisis de datos tradicionales.
 - b) Big Data no desafía la capacidad de las herramientas tradicionales de procesamiento de datos para capturar, almacenar, gestionar y analizar efectivamente la información, dada la evolución constante de las capacidades computacionales y algoritmos de optimización.
 - c) Aunque Big Data puede incluir datos estructurados, su naturaleza se expande al incorporar grandes cantidades de datos no estructurados y semiestructurados, lo que exige el uso de tecnologías especializadas en almacenamiento y procesamiento como Hadoop y sistemas de bases de datos NoSQL.
 - d) Además de su complejidad y diversidad, Big Data introduce desafíos significativos en términos de veracidad y variabilidad al requerir métodos avanzados de limpieza y validación de datos para asegurar la integridad del análisis.
6. La premisa básica del algoritmo de PageRank para clasificar páginas web en los resultados de búsqueda es:
 - a) Los enlaces entrantes a una página web desde otras páginas contribuyen a su importancia y clasificación.
 - b) La antigüedad de la página web es el principal factor para determinar su clasificación.
 - c) La cantidad de visitas que recibe una página web determina su clasificación.
 - d) El contenido y la relevancia de las palabras clave en la página web determinan su posición.
7. En el algoritmo de Indexación Basada en Clasificación Bloqueada (BSBI), ¿cuál es el paso final para crear un índice invertido para la colección completa de los datos?
 - a) Dividir la colección de datos en bloques de tamaño fijo.

- b) Eliminar los términos duplicados de los índices.
 - c) Indexar cada bloque de forma independiente.
 - d) Fusionar los índices invertidos de cada bloque.
8. En un sistema de control de tráfico urbano basado en reglas, ¿cuál de las siguientes reglas sería más efectiva para manejar situaciones de congestión en una intersección?
- a) Si hay pocos vehículos en la intersección, reducir el tiempo de los semáforos en verde.
 - b) Si hay muchos vehículos en la intersección, aumentar el tiempo de los semáforos en verde.
 - c) Si hay un vehículo de emergencia en la intersección, detener todos los demás vehículos.
 - d) Si hay muchos peatones cruzando la intersección, reducir el tiempo de los semáforos en rojo.
9. El algoritmo Hypertext Induced Topic Selection (HITS) intenta buscar nodos especiales. Estos son conocidos como:
- a) Sensibles.
 - b) Centrales.
 - c) Hubs.
 - d) Autoridades.
10. En un sistema donde el conocimiento está orientado a objetos se puede asegurar que:
- a) Los marcos y las bandas son las estructuras utilizadas para la representación del modelo.
 - b) Las funciones de agregación dificultan poder establecer relaciones entre los objetos.
 - c) Se enfatiza la atención a la información de la cual se extrajo el conocimiento.
 - d) No es posible definir especializaciones de los objetos de la vida real dentro del sistema.
11. Una plataforma de comercio electrónico desea mejorar la experiencia del usuario al permitir una navegación más personalizada y contextualizada. Actualmente los usuarios tienen dificultades para encontrar productos relevantes debido a la gran cantidad de opciones disponibles. La empresa está interesada en implementar características de la Web 2.5 y la Web Semántica para abordar este problema. ¿Qué características podrían ayudar para ofrecer una navegación más personalizada y contextualizada?
- a) La implementación de ontologías y metadatos para enriquecer la descripción de productos y mejorar la precisión de las recomendaciones.
 - b) La optimización de la velocidad de carga del sitio web para mejorar la experiencia del usuario y reducir el abandono del carrito de compra.
 - c) La incorporación de los productos en tendencia en el mercado.
 - d) La integración de redes sociales para permitir la recomendación de productos basada en las preferencias de amigos y contactos.
12. La transición de la Web 1.0 a la Web 2.0 se caracterizó principalmente por:
- a) La reducción en el uso de HTML y CSS en el desarrollo de sitios web.
 - b) La disminución de la importancia de los motores de búsqueda en la navegación web.
 - c) El cambio de páginas web estáticas a dinámicas, permitiendo la interacción del usuario y la generación de contenido.
 - d) El aumento en la velocidad de conexión a internet, que permitió una mejor calidad de las páginas web.
13. Analizar una red permite:
- a) Evaluar la calidad del contenido de un sitio web.
 - b) Obtener predicciones exactas de eventos futuros en mercados financieros.
 - c) Detectar posibles tendencias antes de que se conviertan en tendencia.
 - d) Encontrar nodos “sensibles” o críticos para la red.
14. En un grafo una comunidad es:
- a) Un conjunto de nodos altamente conectados que forman un subgrafo completamente independiente del resto de la red.
 - b) Un conjunto de nodos aislados.
 - c) Un conjunto de nodos que no comparten ninguna similitud estructural o funcional entre sí.
 - d) Un conjunto de nodos que están más densamente interconectados entre sí que con los nodos fuera del conjunto.
15. La computación evolutiva:
- a) Utiliza algoritmos para buscar soluciones óptimas en grandes espacios de búsqueda, lo que la hace adecuada para problemas de optimización en la RI.
 - b) Es útil solo para problemas de clasificación de documentos y no para otras tareas de RI en general.
 - c) No es aplicable en la RI debido a la complejidad de los algoritmos evolutivos.
 - d) Solo puede manejar conjuntos de datos pequeños y no es escalable a grandes volúmenes de datos.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Considerando las prácticas éticas y legales en el Web Scraping, se puede asegurar que:
 - a) Web Scraping sobre datos personales sin consentimiento es generalmente aceptado si los datos se utilizan con fines de investigación.
 - b) Es importante revisar y respetar los términos de servicio del sitio web, así como las leyes aplicables de protección de datos y derechos de autor, antes de realizar Web Scraping.
 - c) Web Scraping requiere considerar las políticas de `robots.txt` del sitio web objetivo.
 - d) La extracción de datos mediante Web Scraping siempre es legal, independientemente de las leyes locales sobre derechos de autor y privacidad de datos definidos en los sitios web.
2. Con respecto a WordNet y su contribución a las ontologías y la representación del conocimiento en los sistemas de RI, puede afirmarse que:
 - a) En WordNet, una palabra está asociada a un conjunto de sinónimos (synsets), siendo estas palabras intercambiables en un contexto.
 - b) WordNet diferencia claramente entre los significados de palabras según su uso en diferentes contextos, lo que permite aplicaciones avanzadas en desambiguación semántica más allá de los sistemas de recomendación y RI.
 - c) La integración de WordNet en sistemas de RI permite la expansión de consultas y la mejora de la precisión de los resultados al entender mejor el significado de los términos de búsqueda a través de su contexto semántico y las relaciones entre palabras.
 - d) Una limitación de WordNet en la representación del conocimiento es su enfoque en el idioma inglés, lo que plantea desafíos en la aplicación global y la interoperabilidad con sistemas de información multilingües.
3. Una buena práctica de SEO para mejorar el posicionamiento de un sitio web en los motores de búsqueda es:
 - a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Copiar contenido directamente de otros sitios web populares para aumentar la cantidad de páginas indexadas.
 - c) Llenar el contenido con palabras clave irrelevantes para aumentar la densidad de palabras clave.
 - d) Obtener enlaces de otros sitios web relevantes y de calidad que apunten al sitio.
4. En el contexto de la RI en redes, se puede afirmar que:
 - a) La detección de comunidades ayuda a identificar grupos de nodos altamente conectados entre sí, lo que puede ser útil para comprender la estructura y el contenido de la red.
 - b) La detección de comunidades en una red siempre produce resultados objetivos y consistentes independientemente del algoritmo utilizado.
 - c) La detección de comunidades solo se aplica a redes pequeñas y simples, no a redes grandes y complejas.
 - d) La detección de comunidades no es relevante para la RI en redes, ya que se centra únicamente en la estructura de la red sin considerar el contenido.
5. La afirmación que mejor describe la política de amabilidad en los Web Crawlers es:
 - a) Los Web Crawlers son libres de recopilar datos de cualquier sitio web sin restricciones, independientemente de la cantidad de tráfico que generen.
 - b) La política de amabilidad establece pautas y reglas sobre cómo los crawlers deben interactuar con los sitios web para minimizar la carga del servidor y respetar las directivas de los administradores del sitio.
 - c) La política de amabilidad de los Web Crawlers dicta que los crawlers deben priorizar ciertos tipos de contenido sobre otros, ignorando completamente ciertas páginas web.
 - d) Los crawlers se diseñan para acceder a sitios web sin restricciones y extraer datos de manera agresiva para su indexación.
6. El componente responsable de la gestión de recursos y planificación de tareas en Hadoop es:

- a) MapReduce.
 - b) HDFS.
 - c) Hadoop Common.
 - d) YARN.
7. Considerando las dimensiones y desafíos inherentes a Big Data puede afirmarse, tomando en cuenta las características clave y las implicaciones para su procesamiento y análisis, que:
- a) Big Data se caracteriza principalmente por su pequeño volumen y uniformidad, permitiendo un procesamiento eficiente con mínimas adaptaciones de las herramientas de análisis de datos tradicionales.
 - b) Big Data no desafía la capacidad de las herramientas tradicionales de procesamiento de datos para capturar, almacenar, gestionar y analizar efectivamente la información, dada la evolución constante de las capacidades computacionales y algoritmos de optimización.
 - c) Aunque Big Data puede incluir datos estructurados, su naturaleza se expande al incorporar grandes cantidades de datos no estructurados y semiestructurados, lo que exige el uso de tecnologías especializadas en almacenamiento y procesamiento como Hadoop y sistemas de bases de datos NoSQL.
 - d) Además de su complejidad y diversidad, Big Data introduce desafíos significativos en términos de veracidad y variabilidad al requerir métodos avanzados de limpieza y validación de datos para asegurar la integridad del análisis.
8. En el contexto de la RI en redes sociales se puede afirmar que:
- a) El análisis de redes sociales se centra exclusivamente en la cantidad de seguidores que tiene un usuario en particular para determinar su influencia en la red.
 - b) El análisis de redes sociales no es útil para comprender la difusión de información en una red social específica, para ello se utiliza la medida de centralidad de vector propio.
 - c) El análisis de centralidad de intermediación se utiliza para identificar usuarios que son importantes en una red social debido a su posición como “puentes” entre diferentes grupos de usuarios.
 - d) El análisis de sentimientos se utiliza para determinar la popularidad de una publicación en redes sociales sin tener en cuenta la opinión de los usuarios.
9. En una empresa comercial, se tiene una red donde los nodos corresponden a los empleados y las aristas representan la frecuencia con la que colaboran en las ventas. Se busca mejorar la colaboración entre los empleados para aumentar las ventas totales, por lo que la directiva debe:
- a) No tener en cuenta la centralidad de cercanía de los empleados al diseñar estrategias de colaboración, ya que esto podría introducir complicaciones adicionales en el proceso de trabajo.
 - b) Utilizar la centralidad de cercanía para identificar los empleados menos cercanos a otros en la red y asignarles tareas individuales para evitar posibles conflictos y desacuerdos en el proceso de colaboración.
 - c) No considerar la centralidad de cercanía de los empleados al diseñar estrategias de colaboración, ya que todos los empleados tienen el mismo acceso a la información y los recursos.
 - d) Utilizar la centralidad de cercanía para identificar los empleados más cercanos a otros en la red y promover la colaboración entre ellos, facilitando así la comunicación y el intercambio de conocimientos para mejorar las ventas.
10. La afirmación que mejor refleja el principio subyacente de PageRank, considerando su importancia en la clasificación de los sitios web, es:
- a) El algoritmo de PageRank considera tanto la cantidad como la calidad de los enlaces entrantes, asignando mayor valor a los enlaces provenientes de sitios web considerados como “importantes”.
 - b) PageRank valora más la cantidad de enlaces entrantes a una página web, independientemente de la calidad o relevancia de estos enlaces.
 - c) La efectividad de PageRank se basa exclusivamente en el análisis de las palabras clave contenidas en los enlaces entrantes, sin considerar la estructura de enlace de la Web.
 - d) PageRank opera bajo el supuesto de que los enlaces entrantes y salientes tienen el mismo impacto en la valoración de la relevancia de una página web.
11. En el contexto de la representación del conocimiento basada en herencia, ¿qué caracteriza a la herencia cancelable?
- a) Las propiedades heredadas siempre se mantienen y no pueden anularse.
 - b) La herencia es el resultado del razonamiento no transitivo.
 - c) Las conclusiones no están determinadas y dependen del nodo de interés.
 - d) No existe ambigüedad en las conclusiones obtenidas.
12. En la RI en el contexto de Big Data, se puede asegurar que:

- a) El almacenamiento distribuido es una técnica obsoleta.
 - b) El almacenamiento distribuido centraliza todos los datos en un único servidor para facilitar su gestión y mantenimiento.
 - c) El almacenamiento distribuido reparte los datos en múltiples servidores para mejorar la disponibilidad y la redundancia del sistema.
 - d) El almacenamiento distribuido ofrece ventajas significativas en términos de escalabilidad y rendimiento en comparación con el almacenamiento centralizado.
13. Si una red cumple la propiedad de ser un grafo de mundo pequeño, entonces se conoce que:
- a) La red posee pocos vértices.
 - b) La longitud media del camino entre todo par de vértices es pequeña.
 - c) El número de componentes fuertemente conexas está relacionado con la cantidad de grafos K_n presentes.
 - d) El grafo es un anillo regular de grado 5.
14. Una plataforma de comercio electrónico desea mejorar la experiencia del usuario al permitir una navegación más personalizada y contextualizada. Actualmente los usuarios tienen dificultades para encontrar productos relevantes debido a la gran cantidad de opciones disponibles. La empresa está interesada en implementar características de la Web 2.5 y la Web Semántica para abordar este problema. ¿Qué características podrían ayudar para ofrecer una navegación más personalizada y contextualizada?
- a) La implementación de ontologías y metadatos para enriquecer la descripción de productos y mejorar la precisión de las recomendaciones.
 - b) La optimización de la velocidad de carga del sitio web para mejorar la experiencia del usuario y reducir el abandono del carrito de compra.
 - c) La incorporación de los productos en tendencia en el mercado.
 - d) La integración de redes sociales para permitir la recomendación de productos basada en las preferencias de amigos y contactos.
15. Al “relajar” el concepto de clique en la detección de comunidades se intenta solucionar:
- a) La uniformidad de los nodos en términos de grado.
 - b) La necesidad de datos externos para analizar la red.
 - c) La imposibilidad de conectar nodos distantes.
 - d) El solapamiento y la complejidad computacional.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Una empresa de comercio electrónico necesita mejorar su motor de búsqueda para proporcionar resultados más relevantes a sus usuarios. Actualmente, los resultados de la búsqueda no son precisos y los usuarios a menudo encuentran dificultades para encontrar productos específicos. La empresa está considerando implementar una indexación por conceptos para mejorar la relevancia de los resultados de búsqueda. ¿Qué beneficios podría provocar este cambio?
 - a) Permite adaptarse fácilmente a cambios en el vocabulario y la terminología utilizada en las descripciones de los productos.
 - b) Facilita la visualización de productos al agruparlos por categorías o características comunes.
 - c) Mejora la precisión en la búsqueda de productos relacionados, incluso cuando no coinciden exactamente con los términos de búsqueda del usuario.
 - d) Ayuda a identificar automáticamente relaciones entre productos, lo que puede mejorar las recomendaciones personalizadas a los usuarios.
2. Referente al Web Crawler puede afirmarse que:
 - a) Los hipervínculos encontrados en cada sitio web que no pertenecen al dominio donde fueron encontrados se desechan, puesto que no expande el conjunto de URLs sin visitar.
 - b) Las páginas visitadas no se procesan nunca más.
 - c) No necesita de un conjunto inicial de URLs para recorrer la Web.
 - d) No tiene como objetivo indexar y recopilar información de diferentes sitios web.
3. Una plataforma de comercio electrónico desea mejorar la experiencia del usuario al permitir una navegación más personalizada y contextualizada. Actualmente los usuarios tienen dificultades para encontrar productos relevantes debido a la gran cantidad de opciones disponibles. La empresa está interesada en implementar características de la Web 2.5 y la Web Semántica para abordar este problema. ¿Qué características podrían ayudar para ofrecer una navegación más personalizada y contextualizada?
 - a) La implementación de ontologías y metadatos para enriquecer la descripción de productos y mejorar la precisión de las recomendaciones.
 - b) La optimización de la velocidad de carga del sitio web para mejorar la experiencia del usuario y reducir el abandono del carrito de compra.
 - c) La incorporación de los productos en tendencia en el mercado.
 - d) La integración de redes sociales para permitir la recomendación de productos basada en las preferencias de amigos y contactos.
4. El concepto de *rank sink* en el algoritmo de PageRank representa páginas web:
 - a) Con una baja calidad de contenido y una cantidad insuficiente de enlaces salientes, lo que las hace menos relevantes en los resultados de búsqueda.
 - b) Con distribuciones uniformes de la puntuación de PageRank.
 - c) Con un alto número de enlaces salientes que no reciben enlaces entrantes, lo que puede afectar negativamente su puntuación de PageRank.
 - d) Que tienen una puntuación de PageRank más alta que otras debido a la manipulación de enlaces entrantes y salientes.
5. La centralidad de intermediación de un nodo indica:
 - a) La frecuencia con la que un nodo actúa como puente en el camino más corto entre otros dos nodos.
 - b) El número total de conexiones entrantes y salientes.
 - c) La resistencia del nodo a fallos.
 - d) La cantidad de nodos vecinos directos.
6. ¿Qué estrategia utilizan los Web Crawlers para asegurar un rastreo eficiente y respetuoso de los recursos de los sitios web?
 - a) Extraer únicamente contenido multimedia para reducir la carga en los servidores web.
 - b) Visitar y rastrear todos los enlaces de una página web simultáneamente.
 - c) Ignorar completamente el archivo `robots.txt` de los sitios web.

- d) Seguir las directrices del archivo `robots.txt` y aplicar un retraso entre las solicitudes.
7. La Web 3 se conoce como:
- Web de solo lectura.
 - Internet de las cosas.
 - Web de escritura-lectura.
 - Web semántica.
8. La integración de Hadoop y MapReduce en la RI trae como ventaja:
- La posibilidad de la extracción de información relevante y la generación de resultados significativos de grandes conjuntos de datos.
 - La reducción de los costos operativos a cero.
 - La eliminación de la necesidad de sistemas de bases de datos.
 - La garantía de la privacidad absoluta de los datos procesados.
9. La diferencia entre la indexación por tokens y la indexación por conceptos puede definirse como:
- La indexación por tokens normaliza los datos reduciéndolos a su forma básica, mientras que la indexación por conceptos utiliza un método de ordenación para organizar los términos característicos de los datos.
 - La indexación por tokens asigna un valor numérico a cada término de los datos, mientras que la indexación por conceptos utiliza algoritmos de encriptación para proteger la privacidad de los datos.
 - La indexación por tokens asigna pesos a los términos basados en su importancia relativa, mientras que la indexación por conceptos utiliza un sistema de etiquetado para asociar términos con características generales.
 - La indexación por tokens divide los datos en términos individuales, mientras que la indexación por conceptos agrupa los datos en categorías definidas.
10. Existen varias estrategias consideradas como no recomendables para mejorar el SEO de un sitio web. Dentro de estas estrategias negativas, se encuentran:
- Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - Evitar el uso de las meta etiquetas y descripciones del sitio web para reflejar el contenido de la página.
 - Construir enlaces naturales de sitios web con autoridad y relevancia temática.
- d) Incluir una cantidad excesiva de palabras clave irrelevantes para intentar manipular los rankings de búsqueda.
11. Para la RI, el análisis de las redes puede:
- Ayudar a identificar grupos de interés.
 - Indicar la importancia de una entidad en la transmisión de la información.
 - Revelar patrones de influencia dentro de una comunidad.
 - Ayudar a comprender la conectividad y la accesibilidad entre las entidades.
12. Dentro del ecosistema de Hadoop, el HDFS se caracteriza por:
- La capacidad ilimitada de almacenamiento.
 - La tolerancia a fallos mediante la replicación de datos.
 - El almacenamiento exclusivo para archivos de texto.
 - El modelo de acceso y de escritura de datos en tiempo real.
13. La web actual se enfrenta a problemas como:
- Presencia de una alta calidad en los datos.
 - Heterogeneidad en los datos.
 - Presencia de datos volátiles y distribuidos.
 - Presencia de grandes volúmenes de datos estructurados.
14. La afirmación que mejor describe la política de amabilidad en los Web Crawlers es:
- Los Web Crawlers son libres de recopilar datos de cualquier sitio web sin restricciones, independientemente de la cantidad de tráfico que generen.
 - La política de amabilidad establece pautas y reglas sobre cómo los crawlers deben interactuar con los sitios web para minimizar la carga del servidor y respetar las directivas de los administradores del sitio.
 - La política de amabilidad de los Web Crawlers dicta que los crawlers deben priorizar ciertos tipos de contenido sobre otros, ignorando completamente ciertas páginas web.
 - Los crawlers se diseñan para acceder a sitios web sin restricciones y extraer datos de manera agresiva para su indexación.
15. Implementar índices invertidos en un SRI asegura:
- Incrementar la seguridad de los datos almacenados al dificultar el acceso directo a la información sin el índice correcto.
 - Reducir la cantidad de espacio de almacenamiento necesario al comprimir los datos de los documentos.

c) Acelerar el proceso de búsqueda al permitir búsquedas directas por contenido en lugar de por título.

d) Facilitar la búsqueda sobre los datos que contienen términos específicos al mantener una lista de datos para cada término único.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. En la RI en el contexto de Big Data, se puede asegurar que:
 - a) La indexación distribuida es una técnica obsoleta.
 - b) La indexación distribuida almacena todos los datos en un solo servidor para facilitar su acceso y búsqueda.
 - c) La indexación distribuida divide los datos en múltiples fragmentos que se almacenan en varios nodos para permitir búsquedas paralelas y mejorar la escalabilidad.
 - d) La indexación distribuida no ofrece ventajas en términos de rendimiento y escalabilidad en comparación con la indexación centralizada.
2. El propósito de la política de revisitado en los Web Crawlers es:
 - a) Determinar la frecuencia con la que los crawlers deben volver a visitar una URL específica para mantener la información actualizada en el índice de búsqueda.
 - b) Limitar el acceso de los crawlers a ciertos servidores luego de visitar las páginas alojadas en estos.
 - c) Establecer reglas sobre el tiempo máximo que los crawlers pueden pasar en un sitio web durante cada visita para evitar sobrecargar los servidores.
 - d) Definir la prioridad de rastreo de las URLs, determinando el orden en que los crawlers visitan y procesan cada página web.
3. Dentro del ecosistema de Hadoop, el HDFS se caracteriza por:
 - a) La capacidad ilimitada de almacenamiento.
 - b) La tolerancia a fallos mediante la replicación de datos.
 - c) El almacenamiento exclusivo para archivos de texto.
 - d) El modelo de acceso y de escritura de datos en tiempo real.
4. La diferencia entre la indexación por tokens y la indexación por conceptos puede definirse como:
 - a) La indexación por tokens normaliza los datos reduciéndolos a su forma básica, mientras que la indexación por conceptos utiliza un método de ordenación para organizar los términos característicos de los datos.
 - b) La indexación por tokens asigna un valor numérico a cada término de los datos, mientras que la indexación por conceptos utiliza algoritmos de encriptación para proteger la privacidad de los datos.
 - c) La indexación por tokens asigna pesos a los términos basados en su importancia relativa, mientras que la indexación por conceptos utiliza un sistema de etiquetado para asociar términos con características generales.
 - d) La indexación por tokens divide los datos en términos individuales, mientras que la indexación por conceptos agrupa los datos en categorías definidas.
5. ¿Qué es un 'Uniform Resource Locator (URL) Frontier' en el contexto de Web Crawling?
 - a) Una base de datos que almacena URLs únicas identificadas como recursos en la web.
 - b) Una técnica para filtrar URLs irrelevantes y mejorar la eficiencia del crawling.
 - c) Un protocolo que define cómo se deben formatear las URLs para el crawling.
 - d) Una lista prioritaria de URLs que aún no han sido visitadas por el crawler.
6. Un investigador necesita recopilar datos de múltiples sitios web para un estudio académico, pero se enfrenta a varios desafíos al realizar el proceso de Web Scraping de manera ética y legal. ¿Cuál de las siguientes opciones describe mejor uno de los desafíos asociados al proceso de Web Scraping?
 - a) La dificultad para encontrar herramientas de Web Scraping gratuitas y fiables.
 - b) La disponibilidad limitada de datos en línea que se pueden extraer utilizando técnicas de Web Scraping.
 - c) La necesidad de estar montado sobre un crawler que cumpla con todas las políticas.

- d) La necesidad de comprender la estructura del sitio web y su código HTML para extraer los datos correctamente.
7. Existen varias estrategias consideradas como no recomendables para mejorar el SEO de un sitio web. Dentro de estas estrategias negativas, se encuentran:
- Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - Evitar el uso de las meta etiquetas y descripciones del sitio web para reflejar el contenido de la página.
 - Construir enlaces naturales de sitios web con autoridad y relevancia temática.
 - Incluir una cantidad excesiva de palabras clave irrelevantes para intentar manipular los rankings de búsqueda.
8. En el contexto de la representación del conocimiento basada en herencia, ¿qué caracteriza a la herencia cancelable?
- Las propiedades heredadas siempre se mantienen y no pueden anularse.
 - La herencia es el resultado del razonamiento no transitivo.
 - Las conclusiones no están determinadas y dependen del nodo de interés.
 - No existe ambigüedad en las conclusiones obtenidas.
9. Los Web Crawlers se enfrentan a desafíos constantes. Dentro de ellos se encuentran:
- La falta de acceso a la base de datos del servidor web para extraer información actualizada.
 - La modificación del código y la estructura del sitio web.
 - La incapacidad para interpretar correctamente el lenguaje de programación utilizado en el desarrollo de los sitios web.
 - La dificultad para generar el contenido dinámico en tiempo real.
10. La Web 1.0 se caracteriza por:
- Los propietarios de los sitios web proporcionan contenido de forma periódica.
 - Uso de comunidades virtuales para popularizar los sitios web de noticias.
 - Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
- d) Los sitios web se centran en brindar información en lugar de facilitar la colaboración o participación de los usuarios.
11. En una red de transporte donde cada nodo es una parada de autobús y las aristas representan si existe un carro que pasa por ambos sitios, ¿qué puede mejorar el sistema de transporte?
- Utilizar la centralidad de intermediación para identificar las paradas de transferencia clave y establecer nuevas conexiones entre diferentes líneas de transporte público.
 - Utilizar el grafo inducido de los nodos con mayor valor en la centralidad de grado para aplicar la centralidad de intermediación con el propósito de reforzar las paradas con mayor tráfico.
 - No implementar sistemas de información en tiempo real para los usuarios, ya que pueden aumentar la carga de los trabajadores.
 - No considerar la centralidad de intermediación de las estaciones de transporte público al planificar rutas y horarios, ya que no tiene impacto en las conexiones entre las paradas.
12. Con respecto a WordNet y su contribución a las ontologías y la representación del conocimiento en los sistemas de RI, puede afirmarse que:
- En WordNet, una palabra está asociada a un conjunto de sinónimos (synsets), siendo estas palabras intercambiables en un contexto.
 - WordNet diferencia claramente entre los significados de palabras según su uso en diferentes contextos, lo que permite aplicaciones avanzadas en desambiguación semántica más allá de los sistemas de recomendación y RI.
 - La integración de WordNet en sistemas de RI permite la expansión de consultas y la mejora de la precisión de los resultados al entender mejor el significado de los términos de búsqueda a través de su contexto semántico y las relaciones entre palabras.
 - Una limitación de WordNet en la representación del conocimiento es su enfoque en el idioma inglés, lo que plantea desafíos en la aplicación global y la interoperabilidad con sistemas de información multilingües.
13. En el contexto de la RI en redes sociales se puede afirmar que:
- El análisis de redes sociales se centra exclusivamente en la cantidad de seguidores que tiene un usuario en particular para determinar su influencia en la red.
 - El análisis de redes sociales no es útil para comprender la difusión de información en una red

social específica, para ello se utiliza la medida de centralidad de vector propio.

- c) El análisis de centralidad de intermediación se utiliza para identificar usuarios que son importantes en una red social debido a su posición como “puentes” entre diferentes grupos de usuarios.
 - d) El análisis de sentimientos se utiliza para determinar la popularidad de una publicación en redes sociales sin tener en cuenta la opinión de los usuarios.
14. Sobre el algoritmo de PageRank, visto en clase, se puede afirmar que:
- a) Evalúa la importancia de un sitio web en función de la calidad y cantidad de enlaces entrantes que recibe de otros sitios web.
 - b) Asigna una puntuación baja a los sitios web que contienen muchos enlaces salientes, ya que indica una falta de relevancia.
 - c) Asigna una puntuación alta a los sitios web que tienen un gran número de enlaces entrantes sin tener en cuenta la calidad de esos enlaces.
 - d) Solo tiene en cuenta el contenido en un sitio web para determinar su relevancia en los resultados de búsqueda.
15. Para contribuir positivamente al posicionamiento orgánico de un sitio web en los motores de búsqueda se puede:
- a) Crear contenido relevante y de alta calidad que satisfaga las necesidades de información de los usuarios.
 - b) Obtener enlaces entrantes de otros sitios web de alta autoridad y relevancia temática.
 - c) Incluir una densidad alta de palabras clave para asegurar que el sitio web aparezca en tantas búsquedas como sea posible.
 - d) Mejorar la velocidad de carga del sitio web y asegurar que sea *responsive* y fácil de usar en dispositivos móviles.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Analizar una red permite:
 - a) Evaluar la calidad del contenido de un sitio web.
 - b) Obtener predicciones exactas de eventos futuros en mercados financieros.
 - c) Detectar posibles tendencias antes de que se conviertan en tendencia.
 - d) Encontrar nodos “sensibles” o críticos para la red.
2. En la RI en el contexto de Big Data, se puede asegurar que:
 - a) El almacenamiento distribuido es una técnica obsoleta.
 - b) El almacenamiento distribuido centraliza todos los datos en un único servidor para facilitar su gestión y mantenimiento.
 - c) El almacenamiento distribuido reparte los datos en múltiples servidores para mejorar la disponibilidad y la redundancia del sistema.
 - d) El almacenamiento distribuido ofrece ventajas significativas en términos de escalabilidad y rendimiento en comparación con el almacenamiento centralizado.
3. Los algoritmos para detectar comunidades en una red intentan:
 - a) Buscar subgrafos tal que no incluyan nodos cuya ausencia desconecte al subgrafo.
 - b) Seleccionar aleatoriamente nodos de alto grado y sus vecinos.
 - c) Buscar conjuntos donde cada nodo de un mismo conjunto tenga características similares al resto de los nodos del conjunto.
 - d) Encontrar grupos donde los nodos pertenecientes a los mismos grupos son cercanos bajo cierta métrica y lejanos con respecto a los nodos de otros grupos.
4. En el diseño de un videojuego de roles (RPG) en el que los personajes tienen atributos como salud, fuerza y velocidad, ¿cuál de las siguientes opciones representa mejor una implementación de la representación del conocimiento orientado a objetos?
 - a) Cada personaje se representa como un objeto con propiedades como salud, fuerza y velocidad, y métodos para modificar y consultar estos valores.
 - b) Cada personaje se representa como una función que calcula sus atributos en función de su nivel y experiencia.
 - c) Cada personaje se representa como una matriz de números que almacena sus valores de atributos.
 - d) Cada personaje se representa como una lista de cadenas de texto que describen sus características físicas y habilidades.
5. Dentro del análisis de redes, la centralidad de grado mide la importancia de un nodo basándose en:
 - a) La cantidad de vecinos del nodo.
 - b) El número de aristas que posee el nodo.
 - c) La cantidad de veces que aparece el nodo en el camino mínimo entre cualquier par de nodos.
 - d) El grado del nodo.
6. En un SRI la indexación:
 - a) Mejora la experiencia del usuario.
 - b) Consiste en asociar un identificador único a cada dato almacenado en el sistema.
 - c) Facilita la RI relevante.
 - d) Permite la organización y la categorización de la información.
7. La afirmación que mejor refleja el principio subyacente de PageRank, considerando su importancia en la clasificación de los sitios web, es:
 - a) El algoritmo de PageRank considera tanto la cantidad como la calidad de los enlaces entrantes, asignando mayor valor a los enlaces provenientes de sitios web considerados como “importantes”.
 - b) PageRank valora más la cantidad de enlaces entrantes a una página web, independientemente de la calidad o relevancia de estos enlaces.
 - c) La efectividad de PageRank se basa exclusivamente en el análisis de las palabras clave contenidas en los enlaces entrantes, sin considerar la estructura de enlace de la Web.

- d) PageRank opera bajo el supuesto de que los enlaces entrantes y salientes tienen el mismo impacto en la valoración de la relevancia de una página web.
8. La transición de la Web 1.0 a la Web 2.0 se caracterizó principalmente por:
- La reducción en el uso de HTML y CSS en el desarrollo de sitios web.
 - La disminución de la importancia de los motores de búsqueda en la navegación web.
 - El cambio de páginas web estáticas a dinámicas, permitiendo la interacción del usuario y la generación de contenido.
 - El aumento en la velocidad de conexión a internet, que permitió una mejor calidad de las páginas web.
9. Al realizar la optimización de contenido para SEO debe considerarse:
- Utilizar etiquetas de título y meta descripciones únicas y relevantes para cada página.
 - Seleccionar las palabras clave al azar.
 - Crear contenido valioso y original que satisfaga las necesidades de los usuarios.
 - Incluir palabras clave de manera excesiva en el contenido para mejorar el posicionamiento.
10. Se puede afirmar que:
- Uno de los desafíos en el procesamiento de Big Data es la capacidad de gestionar y analizar datos provenientes de diversas fuentes y en diferentes formatos de manera eficiente.
 - MapReduce es un modelo de procesamiento distribuido utilizado para trabajar con grandes volúmenes de datos.
 - Los SRI pueden beneficiarse de MapReduce para mejorar la RI relevante.
 - El término “Big Data” se refiere exclusivamente al volumen de datos que una organización maneja, sin tener en cuenta la velocidad, la variedad y la veracidad de los datos.
11. El posicionamiento de un sitio web en los motores de búsqueda puede ser afectado por:
- Mantener una estructura de URL clara y coherente.
 - Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - Contenido sin valor en el sitio web.
 - Obtener enlaces de sitios web irrelevantes y de baja calidad.
12. Sobre WordNet y su aplicación en el procesamiento del lenguaje natural, se puede afirmar que:
- Aunque WordNet es una herramienta valiosa en el NLP, su estructura no incluye información sobre la frecuencia de uso de las palabras en el lenguaje natural.
 - El diseño de WordNet facilita su integración en aplicaciones multilingües de NLP, aunque su desarrollo original se centró en el inglés.
 - Los synsets facilitan la identificación de relaciones semánticas entre palabras, como la hiperonimia y la meronimia, enriqueciendo tareas de NLP.
 - WordNet proporciona una base para la desambiguación semántica al agrupar palabras con significados similares en synsets.
13. Dentro del ecosistema de Hadoop, el HDFS se caracteriza por:
- La capacidad ilimitada de almacenamiento.
 - La tolerancia a fallos mediante la replicación de datos.
 - El almacenamiento exclusivo para archivos de texto.
 - El modelo de acceso y de escritura de datos en tiempo real.
14. El algoritmo Hypertext Induced Topic Selection (HITS) intenta buscar nodos especiales. Estos son conocidos como:
- Sensibles.
 - Centrales.
 - Hubs.
 - Autoridades.
15. Una plataforma de comercio electrónico desea mejorar la experiencia del usuario al permitir una navegación más personalizada y contextualizada. Actualmente los usuarios tienen dificultades para encontrar productos relevantes debido a la gran cantidad de opciones disponibles. La empresa está interesada en implementar características de la Web 2.5 y la Web Semántica para abordar este problema. ¿Qué características podrían ayudar para ofrecer una navegación más personalizada y contextualizada?
- La implementación de ontologías y metadatos para enriquecer la descripción de productos y mejorar la precisión de las recomendaciones.
 - La optimización de la velocidad de carga del sitio web para mejorar la experiencia del usuario y reducir el abandono del carrito de compra.
 - La incorporación de los productos en tendencia en el mercado.
 - La integración de redes sociales para permitir la recomendación de productos basada en las preferencias de amigos y contactos.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Los algoritmos para detectar comunidades en una red intentan:
 - a) Buscar subgrafos tal que no incluyan nodos cuya ausencia desconecte al subgrafo.
 - b) Seleccionar aleatoriamente nodos de alto grado y sus vecinos.
 - c) Buscar conjuntos donde cada nodo de un mismo conjunto tenga características similares al resto de los nodos del conjunto.
 - d) Encontrar grupos donde los nodos pertenecientes a los mismos grupos son cercanos bajo cierta métrica y lejanos con respecto a los nodos de otros grupos.
2. Considerando las dimensiones y desafíos inherentes a Big Data puede afirmarse, tomando en cuenta las características clave y las implicaciones para su procesamiento y análisis, que:
 - a) Big Data se caracteriza principalmente por su pequeño volumen y uniformidad, permitiendo un procesamiento eficiente con mínimas adaptaciones de las herramientas de análisis de datos tradicionales.
 - b) Big Data no desafía la capacidad de las herramientas tradicionales de procesamiento de datos para capturar, almacenar, gestionar y analizar efectivamente la información, dada la evolución constante de las capacidades computacionales y algoritmos de optimización.
 - c) Aunque Big Data puede incluir datos estructurados, su naturaleza se expande al incorporar grandes cantidades de datos no estructurados y semiestructurados, lo que exige el uso de tecnologías especializadas en almacenamiento y procesamiento como Hadoop y sistemas de bases de datos NoSQL.
 - d) Además de su complejidad y diversidad, Big Data introduce desafíos significativos en términos de veracidad y variabilidad al requerir métodos avanzados de limpieza y validación de datos para asegurar la integridad del análisis.
3. Considerando las prácticas éticas y legales en el Web Scraping, se puede asegurar que:
 - a) Web Scraping sobre datos personales sin consentimiento es generalmente aceptado si los datos se utilizan con fines de investigación.
 - b) Es importante revisar y respetar los términos de servicio del sitio web, así como las leyes aplicables de protección de datos y derechos de autor, antes de realizar Web Scraping.
 - c) Web Scraping requiere considerar las políticas de `robots.txt` del sitio web objetivo.
 - d) La extracción de datos mediante Web Scraping siempre es legal, independientemente de las leyes locales sobre derechos de autor y privacidad de datos definidos en los sitios web.
4. La política de ordenación de URLs en los Web Crawlers tiene como aspecto fundamental:
 - a) Limitar el acceso de los crawlers a ciertas secciones de un sitio web, evitando el rastreo de URLs consideradas menos importantes o sensibles.
 - b) Establecer la prioridad de rastreo de las URLs, determinando el orden en que los crawlers visitan y procesan cada página web.
 - c) Definir la estructura de la URL de destino, asegurando que estén ordenadas alfabéticamente para facilitar la navegación y la indexación.
 - d) Determinar la forma en que los crawlers asignan un valor de relevancia a cada URL para clasificarlas en el índice de búsqueda.
5. No se considera como técnica para detectar comunidades en una red:
 - a) Analizar la mutualidad de los enlaces.
 - b) Usar el agrupamiento jerárquico.
 - c) Utilizar el algoritmo de K-Means.
 - d) Encontrar cliques de vértices de grado par.
6. En el contexto de la RI en redes, se puede afirmar que:
 - a) La detección de comunidades ayuda a identificar grupos de nodos altamente conectados entre sí, lo que puede ser útil para comprender la estructura y el contenido de la red.

- b) La detección de comunidades en una red siempre produce resultados objetivos y consistentes independientemente del algoritmo utilizado.
 - c) La detección de comunidades solo se aplica a redes pequeñas y simples, no a redes grandes y complejas.
 - d) La detección de comunidades no es relevante para la RI en redes, ya que se centra únicamente en la estructura de la red sin considerar el contenido.
7. Se puede afirmar que:
- a) Uno de los desafíos en el procesamiento de Big Data es la capacidad de gestionar y analizar datos provenientes de diversas fuentes y en diferentes formatos de manera eficiente.
 - b) MapReduce es un modelo de procesamiento distribuido utilizado para trabajar con grandes volúmenes de datos.
 - c) Los SRI pueden beneficiarse de MapReduce para mejorar la RI relevante.
 - d) El término “Big Data” se refiere exclusivamente al volumen de datos que una organización maneja, sin tener en cuenta la velocidad, la variedad y la veracidad de los datos.
8. Sobre WordNet y su aplicación en el procesamiento del lenguaje natural, se puede afirmar que:
- a) Aunque WordNet es una herramienta valiosa en el NLP, su estructura no incluye información sobre la frecuencia de uso de las palabras en el lenguaje natural.
 - b) El diseño de WordNet facilita su integración en aplicaciones multilingües de NLP, aunque su desarrollo original se centró en el inglés.
 - c) Los synsets facilitan la identificación de relaciones semánticas entre palabras, como la hiperonimia y la meronimia, enriqueciendo tareas de NLP.
 - d) WordNet proporciona una base para la desambiguación semántica al agrupar palabras con significados similares en synsets.
9. Existen varias estrategias consideradas como no recomendables para mejorar el SEO de un sitio web. Dentro de estas estrategias negativas, se encuentran:
- a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Evitar el uso de las meta etiquetas y descripciones del sitio web para reflejar el contenido de la página.
 - c) Construir enlaces naturales de sitios web con autoridad y relevancia temática.
 - d) Incluir una cantidad excesiva de palabras clave irrelevantes para intentar manipular los rankings de búsqueda.
10. La web actual se enfrenta a problemas como:
- a) Presencia de una alta calidad en los datos.
 - b) Heterogeneidad en los datos.
 - c) Presencia de datos volátiles y distribuidos.
 - d) Presencia de grandes volúmenes de datos estructurados.
11. La premisa básica del algoritmo de PageRank para clasificar páginas web en los resultados de búsqueda es:
- a) Los enlaces entrantes a una página web desde otras páginas contribuyen a su importancia y clasificación.
 - b) La antigüedad de la página web es el principal factor para determinar su clasificación.
 - c) La cantidad de visitas que recibe una página web determina su clasificación.
 - d) El contenido y la relevancia de las palabras clave en la página web determinan su posición.
12. ¿Cuál de las siguientes opciones describe mejor la diferencia clave entre Web Crawling y Web Scraping?
- a) Web Scraping implica el análisis de la estructura y el contenido de las páginas web para extraer datos, mientras que Web Crawling se refiere a la descarga y almacenamiento de páginas web completas.
 - b) Web Scraping es más eficaz para rastrear e indexar contenido web para motores de búsqueda, mientras que Web Crawling se utiliza principalmente para la extracción de datos en proyectos de investigación.
 - c) Web Crawling se centra en la exploración y recopilación de enlaces de múltiples sitios web, mientras que Web Scraping se enfoca en la extracción específica de datos de páginas web individuales.
 - d) Web Crawling se realiza utilizando herramientas de automatización como Selenium WebDriver, mientras que Web Scraping se lleva a cabo mediante el análisis de HTML y CSS.
13. En el contexto de la representación del conocimiento basada en herencia, ¿qué caracteriza a la herencia cancelable?
- a) Las propiedades heredadas siempre se mantienen y no pueden anularse.
 - b) La herencia es el resultado del razonamiento no transitivo.

- c) Las conclusiones no están determinadas y dependen del nodo de interés.
 - d) No existe ambigüedad en las conclusiones obtenidas.
14. Una buena práctica de SEO para mejorar el posicionamiento de un sitio web en los motores de búsqueda es:
- a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Copiar contenido directamente de otros sitios web populares para aumentar la cantidad de páginas indexadas.
 - c) Llenar el contenido con palabras clave irrelevantes para aumentar la densidad de palabras clave.
15. En un sistema donde el conocimiento está definido a partir de reglas se puede asegurar que:
- a) El nivel de especificidad de las reglas está limitado.
 - b) La representación de conocimiento está basada en la lógica proposicional.
 - c) El razonamiento se activa en cadena hacia delante.
 - d) El orden en que se definen las reglas no altera el razonamiento del sistema.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Al realizar la optimización de contenido para SEO debe considerarse:
 - a) Utilizar etiquetas de título y meta descripciones únicas y relevantes para cada página.
 - b) Seleccionar las palabras clave al azar.
 - c) Crear contenido valioso y original que satisfaga las necesidades de los usuarios.
 - d) Incluir palabras clave de manera excesiva en el contenido para mejorar el posicionamiento.
2. ¿Qué algoritmos permiten obtener información de una red?
 - a) Índices de centralidad.
 - b) Detección de comunidades.
 - c) Hypertext Induced Topic Selection (HITS).
 - d) PageRank.
3. Para transformar el contenido no estructurado de las páginas web en datos estructurados, el Web Scraping:
 - a) Convierte de forma automática imágenes a texto.
 - b) Interpreta el código JavaScript en tiempo real.
 - c) Extrae información basada en patrones de HTML/CSS.
 - d) Analiza los protocolos de red.
4. Los algoritmos para detectar comunidades en una red intentan:
 - a) Buscar subgrafos tal que no incluyan nodos cuya ausencia desconecte al subgrafo.
 - b) Seleccionar aleatoriamente nodos de alto grado y sus vecinos.
 - c) Buscar conjuntos donde cada nodo de un mismo conjunto tenga características similares al resto de los nodos del conjunto.
 - d) Encontrar grupos donde los nodos pertenecientes a los mismos grupos son cercanos bajo cierta métrica y lejanos con respecto a los nodos de otros grupos.
5. En el modelo de representación del conocimiento basado en herencia se puede asegurar que:
 - a) La cantidad de padres que puede tener un nodo no es mayor que 4.
 - b) Solo se usa en entornos referentes a la biología.
 - c) El razonamiento deducido está respaldado por al menos un camino dentro del grafo.
 - d) Las conclusiones pueden ser canceladas si el grafo es ambiguo.
6. Un investigador necesita recopilar datos de múltiples sitios web para un estudio académico, pero se enfrenta a varios desafíos al realizar el proceso de Web Scraping de manera ética y legal. ¿Cuál de las siguientes opciones describe mejor uno de los desafíos asociados al proceso de Web Scraping?
 - a) La dificultad para encontrar herramientas de Web Scraping gratuitas y fiables.
 - b) La disponibilidad limitada de datos en línea que se pueden extraer utilizando técnicas de Web Scraping.
 - c) La necesidad de estar montado sobre un scrawler que cumpla con todas las políticas.
 - d) La necesidad de comprender la estructura del sitio web y su código HTML para extraer los datos correctamente.
7. La premisa básica del algoritmo de PageRank para clasificar páginas web en los resultados de búsqueda es:
 - a) Los enlaces entrantes a una página web desde otras páginas contribuyen a su importancia y clasificación.
 - b) La antigüedad de la página web es el principal factor para determinar su clasificación.
 - c) La cantidad de visitas que recibe una página web determina su clasificación.
 - d) El contenido y la relevancia de las palabras clave en la página web determinan su posición.
8. En una red de transporte donde cada nodo es una parada de autobús y las aristas representan si existe un carro que pasa por ambos sitios, ¿qué puede mejorar el sistema de transporte?
 - a) Utilizar la centralidad de intermediación para identificar las paradas de transferencia clave y establecer nuevas conexiones entre diferentes líneas de transporte público.

- b) Utilizar el grafo inducido de los nodos con mayor valor en la centralidad de grado para aplicar la centralidad de intermediación con el propósito de reforzar las paradas con mayor tráfico.
- c) No implementar sistemas de información en tiempo real para los usuarios, ya que pueden aumentar la carga de los trabajadores.
- d) No considerar la centralidad de intermediación de las estaciones de transporte público al planificar rutas y horarios, ya que no tiene impacto en las conexiones entre las paradas.
9. En el contexto de la RI en redes, se puede afirmar que:
- a) La detección de comunidades ayuda a identificar grupos de nodos altamente conectados entre sí, lo que puede ser útil para comprender la estructura y el contenido de la red.
- b) La detección de comunidades en una red siempre produce resultados objetivos y consistentes independientemente del algoritmo utilizado.
- c) La detección de comunidades solo se aplica a redes pequeñas y simples, no a redes grandes y complejas.
- d) La detección de comunidades no es relevante para la RI en redes, ya que se centra únicamente en la estructura de la red sin considerar el contenido.
10. En la RI en el contexto de Big Data, se puede asegurar que:
- a) La indexación distribuida es una técnica obsoleta.
- b) La indexación distribuida almacena todos los datos en un solo servidor para facilitar su acceso y búsqueda.
- c) La indexación distribuida divide los datos en múltiples fragmentos que se almacenan en varios nodos para permitir búsquedas paralelas y mejorar la escalabilidad.
- d) La indexación distribuida no ofrece ventajas en términos de rendimiento y escalabilidad en comparación con la indexación centralizada.
11. Si una red cumple la propiedad de ser un grafo de mundo pequeño, entonces se conoce que:
- a) La red posee pocos vértices.
- b) La longitud media del camino entre todo par de vértices es pequeña.
- c) El número de componentes fuertemente conexas está relacionado con la cantidad de grafos K_n presentes.
- d) El grafo es un anillo regular de grado 5.
12. Dentro del ecosistema de Hadoop, el HDFS se caracteriza por:
- a) La capacidad ilimitada de almacenamiento.
- b) La tolerancia a fallos mediante la replicación de datos.
- c) El almacenamiento exclusivo para archivos de texto.
- d) El modelo de acceso y de escritura de datos en tiempo real.
13. El conjunto frontera de URLs en un Web Crawler:
- a) Delimita los sitios web que visitará en futuras iteraciones del proceso.
- b) Indica las secciones que no pueden ser visitadas de cada sitio web.
- c) Almacena hipervínculos que pertenecen al mismo dominio del conjunto semilla de URLs con que inició el crawler.
- d) Es similar a un conjunto de URLs que esperan ser visitadas.
14. Un sistema cuenta con la siguiente información:
- Juan pertenece a la marina.
 - Juan es capellán (sacerdote encargado de una tarea específica fuera de la parroquia).
 - Los infantes de la marina suelen ser bebedores de cerveza.
 - Un capellán no suele ser bebedor de cerveza.
 - Un bebedor de cerveza suele tener sobrepeso.
 - Por lo general, un infante de la marina está en buena condición física.
- Si se representa la información en una red de herencia se puede concluir que:
- a) Existen dos razonamientos cancelables.
- b) Juan tiene sobrepeso.
- c) No se puede asegurar que Juan sea capellán.
- d) No se puede asegurar que Juan esté en buena condición física.
15. En el diseño de un videojuego de roles (RPG) en el que los personajes tienen atributos como salud, fuerza y velocidad, ¿cuál de las siguientes opciones representa mejor una implementación de la representación del conocimiento orientado a objetos?
- a) Cada personaje se representa como un objeto con propiedades como salud, fuerza y velocidad, y métodos para modificar y consultar estos valores.
- b) Cada personaje se representa como una función que calcula sus atributos en función de su nivel y experiencia.
- c) Cada personaje se representa como una matriz de números que almacena sus valores de atributos.
- d) Cada personaje se representa como una lista de cadenas de texto que describen sus características físicas y habilidades.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Los Web Crawlers se enfrentan a desafíos constantes. Dentro de ellos se encuentran:
 - a) La falta de acceso a la base de datos del servidor web para extraer información actualizada.
 - b) La modificación del código y la estructura del sitio web.
 - c) La incapacidad para interpretar correctamente el lenguaje de programación utilizado en el desarrollo de los sitios web.
 - d) La dificultad para generar el contenido dinámico en tiempo real.
2. Para transformar el contenido no estructurado de las páginas web en datos estructurados, el Web Scraping:
 - a) Convierte de forma automática imágenes a texto.
 - b) Interpreta el código JavaScript en tiempo real.
 - c) Extrae información basada en patrones de HTML/CSS.
 - d) Analiza los protocolos de red.
3. Para la RI, el análisis de las redes puede:
 - a) Ayudar a identificar grupos de interés.
 - b) Indicar la importancia de una entidad en la transmisión de la información.
 - c) Revelar patrones de influencia dentro de una comunidad.
 - d) Ayudar a comprender la conectividad y la accesibilidad entre las entidades.
4. Dentro del ecosistema de Hadoop, el HDFS se caracteriza por:
 - a) La capacidad ilimitada de almacenamiento.
 - b) La tolerancia a fallos mediante la replicación de datos.
 - c) El almacenamiento exclusivo para archivos de texto.
 - d) El modelo de acceso y de escritura de datos en tiempo real.
5. Se tiene un grafo donde cada nodo es un personaje de cierto libro de cuentos y la existencia de las aristas está definida si dos personajes aparecen en el mismo cuento. Se puede asegurar que:
 - a) La centralidad de grado es útil para identificar los personajes con más conexiones dentro de la red, lo que puede indicar su importancia en el libro.
 - b) La centralidad de cercanía indica el grado de conexión de cada personaje con el resto de los personajes de los cuentos del libro.
 - c) La centralidad de intermediación identifica a los personajes que actúan como conectores entre personajes de cuentos distintos.
 - d) La centralidad de cercanía ofrece una relación entre la cantidad de vecinos de un nodo con respecto a la longitud máxima de un camino dentro del grafo partiendo del nodo en cuestión.
6. Un sistema cuenta con la siguiente información:
 - Juan pertenece a la marina.
 - Juan es capellán (sacerdote encargado de una tarea específica fuera de la parroquia).
 - Los infantes de la marina suelen ser bebedores de cerveza.
 - Un capellán no suele ser bebedor de cerveza.
 - Un bebedor de cerveza suele tener sobrepeso.
 - Por lo general, un infante de la marina está en buena condición física.

Si se representa la información en una red de herencia se puede concluir que:

 - a) Existen dos razonamientos cancelables.
 - b) Juan tiene sobrepeso.
 - c) No se puede asegurar que Juan sea capellán.
 - d) No se puede asegurar que Juan esté en buena condición física.
7. Si una red cumple la propiedad de ser un grafo de mundo pequeño, entonces se conoce que:
 - a) La red posee pocos vértices.
 - b) La longitud media del camino entre todo par de vértices es pequeña.
 - c) El número de componentes fuertemente conexas está relacionado con la cantidad de grafos K_n presentes.
 - d) El grafo es un anillo regular de grado 5.

8. El posicionamiento de un sitio web en los motores de búsqueda puede ser afectado por:
 - a) Mantener una estructura de URL clara y coherente.
 - b) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - c) Contenido sin valor en el sitio web.
 - d) Obtener enlaces de sitios web irrelevantes y de baja calidad.
9. Se puede afirmar que:
 - a) La cantidad de conexiones de un nodo siempre indica su influencia en la red.
 - b) Todas las relaciones en una red tienen la misma importancia para el análisis.
 - c) No es posible identificar subgrupos dentro de una red utilizando análisis de redes.
 - d) El tamaño de una red es siempre indicativo de su efectividad en la transmisión de información.
10. En un sistema de control de tráfico urbano basado en reglas, ¿cuál de las siguientes reglas sería más efectiva para manejar situaciones de congestión en una intersección?
 - a) Si hay pocos vehículos en la intersección, reducir el tiempo de los semáforos en verde.
 - b) Si hay muchos vehículos en la intersección, aumentar el tiempo de los semáforos en verde.
 - c) Si hay un vehículo de emergencia en la intersección, detener todos los demás vehículos.
 - d) Si hay muchos peatones cruzando la intersección, reducir el tiempo de los semáforos en rojo.
11. Si se tiene el conjunto de páginas interconectadas $\{A \rightarrow B, C; B \rightarrow C; C \rightarrow A\}$, entonces la página con valor más alto de PageRank es:
 - a) B.
 - b) C.
 - c) A y B.
 - d) A.
12. Existen varias estrategias consideradas como no recomendables para mejorar el SEO de un sitio web. Dentro de estas estrategias negativas, se encuentran:
 - a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Evitar el uso de las meta etiquetas y descripciones del sitio web para reflejar el contenido de la página.
 - c) Construir enlaces naturales de sitios web con autoridad y relevancia temática.
 - d) Incluir una cantidad excesiva de palabras clave irrelevantes para intentar manipular los rankings de búsqueda.
13. Se tiene una red de ingredientes donde cada uno representa un nodo y las aristas simbolizan que los ingredientes forman parte de una misma receta. Se busca mejorar la experiencia culinaria mediante la elaboración de combinaciones de ingredientes más interesantes y creativas, para lo cual se debe:
 - a) No tener en cuenta la centralidad de grado de los ingredientes para no darle mayor importancia a los ingredientes más comunes.
 - b) No considerar la centralidad de grado de los ingredientes, ya que todas las combinaciones de ingredientes son igualmente válidas.
 - c) Utilizar la centralidad de grado para identificar los ingredientes más populares en la red y crear combinaciones que incluyan una variedad de ingredientes menos comunes.
 - d) Utilizar la centralidad de grado para identificar los ingredientes menos conectados en la red y tomarlos en cuenta para su inclusión en las combinaciones.
14. Sobre WordNet y su aplicación en el procesamiento del lenguaje natural, se puede afirmar que:
 - a) Aunque WordNet es una herramienta valiosa en el NLP, su estructura no incluye información sobre la frecuencia de uso de las palabras en el lenguaje natural.
 - b) El diseño de WordNet facilita su integración en aplicaciones multilingües de NLP, aunque su desarrollo original se centró en el inglés.
 - c) Los synsets facilitan la identificación de relaciones semánticas entre palabras, como la hiperonimia y la meronimia, enriqueciendo tareas de NLP.
 - d) WordNet proporciona una base para la desambiguación semántica al agrupar palabras con significados similares en synsets.
15. El componente responsable de la gestión de recursos y planificación de tareas en Hadoop es:
 - a) MapReduce.
 - b) HDFS.
 - c) Hadoop Common.
 - d) YARN.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Sobre el algoritmo de PageRank, visto en clase, se puede afirmar que:
 - a) Evalúa la importancia de un sitio web en función de la calidad y cantidad de enlaces entrantes que recibe de otros sitios web.
 - b) Asigna una puntuación baja a los sitios web que contienen muchos enlaces salientes, ya que indica una falta de relevancia.
 - c) Asigna una puntuación alta a los sitios web que tienen un gran número de enlaces entrantes sin tener en cuenta la calidad de esos enlaces.
 - d) Solo tiene en cuenta el contenido en un sitio web para determinar su relevancia en los resultados de búsqueda.
2. Para contribuir positivamente al posicionamiento orgánico de un sitio web en los motores de búsqueda se puede:
 - a) Crear contenido relevante y de alta calidad que satisfaga las necesidades de información de los usuarios.
 - b) Obtener enlaces entrantes de otros sitios web de alta autoridad y relevancia temática.
 - c) Incluir una densidad alta de palabras clave para asegurar que el sitio web aparezca en tantas búsquedas como sea posible.
 - d) Mejorar la velocidad de carga del sitio web y asegurar que sea *responsive* y fácil de usar en dispositivos móviles.
3. Los Web Crawlers se enfrentan a desafíos constantes. Dentro de ellos se encuentran:
 - a) La falta de acceso a la base de datos del servidor web para extraer información actualizada.
 - b) La modificación del código y la estructura del sitio web.
 - c) La incapacidad para interpretar correctamente el lenguaje de programación utilizado en el desarrollo de los sitios web.
 - d) La dificultad para generar el contenido dinámico en tiempo real.
4. En el contexto del procesamiento de grandes conjuntos de datos, se puede asegurar que:
 - a) MapReduce no es adecuado para el procesamiento de datos no estructurados.
 - b) MapReduce solo puede manejar pequeñas cantidades de datos y no escala bien a grandes conjuntos de datos.
 - c) MapReduce es un enfoque para procesar datos de forma secuencial en un solo servidor para evitar problemas de concurrencia.
 - d) MapReduce divide una tarea en múltiples pasos de map y reduce que se ejecutan de forma secuencial en diferentes servidores para mejorar el rendimiento y la escalabilidad.
5. En el algoritmo de Indexación Basada en Clasificación Bloqueada (BSBI), ¿cuál es el paso final para crear un índice invertido para la colección completa de los datos?
 - a) Dividir la colección de datos en bloques de tamaño fijo.
 - b) Eliminar los términos duplicados de los índices.
 - c) Indexar cada bloque de forma independiente.
 - d) Fusionar los índices invertidos de cada bloque.
6. Los algoritmos para detectar comunidades en una red intentan:
 - a) Buscar subgrafos tal que no incluyan nodos cuya ausencia desconecte al subgrafo.
 - b) Seleccionar aleatoriamente nodos de alto grado y sus vecinos.
 - c) Buscar conjuntos donde cada nodo de un mismo conjunto tenga características similares al resto de los nodos del conjunto.
 - d) Encontrar grupos donde los nodos pertenecientes a los mismos grupos son cercanos bajo cierta métrica y lejanos con respecto a los nodos de otros grupos.
7. Un SRI es capaz de:
 - a) No necesitar de ningún almacenamiento externo para alojar los índices de los datos.
 - b) Reducir el tiempo de indexación de los datos si utiliza servidores distribuidos dentro de la red para que cada uno ejecute la indexación del mismo conjunto de datos.
 - c) Generar índices invertidos de manera óptima sin considerar el contexto.

- d) Crear los índices asociados a los datos sin tener que analizar cada dato.
8. En un sistema de control de tráfico urbano basado en reglas, ¿cuál de las siguientes reglas sería más efectiva para manejar situaciones de congestión en una intersección?
- Si hay pocos vehículos en la intersección, reducir el tiempo de los semáforos en verde.
 - Si hay muchos vehículos en la intersección, aumentar el tiempo de los semáforos en verde.
 - Si hay un vehículo de emergencia en la intersección, detener todos los demás vehículos.
 - Si hay muchos peatones cruzando la intersección, reducir el tiempo de los semáforos en rojo.
9. En la RI en el contexto de Big Data, se puede asegurar que:
- La indexación distribuida es una técnica obsoleta.
 - La indexación distribuida almacena todos los datos en un solo servidor para facilitar su acceso y búsqueda.
 - La indexación distribuida divide los datos en múltiples fragmentos que se almacenan en varios nodos para permitir búsquedas paralelas y mejorar la escalabilidad.
 - La indexación distribuida no ofrece ventajas en términos de rendimiento y escalabilidad en comparación con la indexación centralizada.
10. Con respecto a WordNet y su contribución a las ontologías y la representación del conocimiento en los sistemas de RI, puede afirmarse que:
- En WordNet, una palabra está asociada a un conjunto de sinónimos (synsets), siendo estas palabras intercambiables en un contexto.
 - WordNet diferencia claramente entre los significados de palabras según su uso en diferentes contextos, lo que permite aplicaciones avanzadas en desambiguación semántica más allá de los sistemas de recomendación y RI.
 - La integración de WordNet en sistemas de RI permite la expansión de consultas y la mejora de la precisión de los resultados al entender mejor el significado de los términos de búsqueda a través de su contexto semántico y las relaciones entre palabras.
 - Una limitación de WordNet en la representación del conocimiento es su enfoque en el idioma inglés, lo que plantea desafíos en la aplicación global y la interoperabilidad con sistemas de información multilingües.
11. ¿Cuál de las siguientes opciones describe mejor la diferencia clave entre Web Crawling y Web Scraping?
- Web Scraping implica el análisis de la estructura y el contenido de las páginas web para extraer datos, mientras que Web Crawling se refiere a la descarga y almacenamiento de páginas web completas.
 - Web Scraping es más eficaz para rastrear e indexar contenido web para motores de búsqueda, mientras que Web Crawling se utiliza principalmente para la extracción de datos en proyectos de investigación.
 - Web Crawling se centra en la exploración y recopilación de enlaces de múltiples sitios web, mientras que Web Scraping se enfoca en la extracción específica de datos de páginas web individuales.
 - Web Crawling se realiza utilizando herramientas de automatización como Selenium WebDriver, mientras que Web Scraping se lleva a cabo mediante el análisis de HTML y CSS.
12. Se tiene una red de ingredientes donde cada uno representa un nodo y las aristas simbolizan que los ingredientes forman parte de una misma receta. Se busca mejorar la experiencia culinaria mediante la elaboración de combinaciones de ingredientes más interesantes y creativas, para lo cual se debe:
- No tener en cuenta la centralidad de grado de los ingredientes para no darle mayor importancia a los ingredientes más comunes.
 - No considerar la centralidad de grado de los ingredientes, ya que todas las combinaciones de ingredientes son igualmente válidas.
 - Utilizar la centralidad de grado para identificar los ingredientes más populares en la red y crear combinaciones que incluyan una variedad de ingredientes menos comunes.
 - Utilizar la centralidad de grado para identificar los ingredientes menos conectados en la red y tomarlos en cuenta para su inclusión en las combinaciones.
13. En una empresa comercial, se tiene una red donde los nodos corresponden a los empleados y las aristas representan la frecuencia con la que colaboran en las ventas. Se busca mejorar la colaboración entre los empleados para aumentar las ventas totales, por lo que la directiva debe:
- No tener en cuenta la centralidad de cercanía de los empleados al diseñar estrategias de colaboración, ya que esto podría introducir complicaciones adicionales en el proceso de trabajo.

- b) Utilizar la centralidad de cercanía para identificar los empleados menos cercanos a otros en la red y asignarles tareas individuales para evitar posibles conflictos y desacuerdos en el proceso de colaboración.
 - c) No considerar la centralidad de cercanía de los empleados al diseñar estrategias de colaboración, ya que todos los empleados tienen el mismo acceso a la información y los recursos.
 - d) Utilizar la centralidad de cercanía para identificar los empleados más cercanos a otros en la red y promover la colaboración entre ellos, facilitando así la comunicación y el intercambio de conocimientos para mejorar las ventas.
14. En el contexto de la RI en redes, se puede afirmar que:
- a) La detección de comunidades ayuda a identificar grupos de nodos altamente conectados entre sí, lo que puede ser útil para comprender la estructura y el contenido de la red.
 - b) La detección de comunidades en una red siempre produce resultados objetivos y consistentes independientemente del algoritmo utilizado.
 - c) La detección de comunidades solo se aplica a redes pequeñas y simples, no a redes grandes y complejas.
 - d) La detección de comunidades no es relevante para la RI en redes, ya que se centra únicamente en la estructura de la red sin considerar el contenido.
15. En un sistema donde el conocimiento está definido a partir de reglas se puede asegurar que:
- a) El nivel de especificidad de las reglas está limitado.
 - b) La representación de conocimiento está basada en la lógica proposicional.
 - c) El razonamiento se activa en cadena hacia delante.
 - d) El orden en que se definen las reglas no altera el razonamiento del sistema.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. En el contexto de la RI en redes sociales se puede afirmar que:
 - a) El análisis de redes sociales se centra exclusivamente en la cantidad de seguidores que tiene un usuario en particular para determinar su influencia en la red.
 - b) El análisis de redes sociales no es útil para comprender la difusión de información en una red social específica, para ello se utiliza la medida de centralidad de vector propio.
 - c) El análisis de centralidad de intermediación se utiliza para identificar usuarios que son importantes en una red social debido a su posición como “puentes” entre diferentes grupos de usuarios.
 - d) El análisis de sentimientos se utiliza para determinar la popularidad de una publicación en redes sociales sin tener en cuenta la opinión de los usuarios.
2. Un sistema cuenta con la siguiente información:
 - Juan pertenece a la marina.
 - Juan es capellán (sacerdote encargado de una tarea específica fuera de la parroquia).
 - Los infantes de la marina suelen ser bebedores de cerveza.
 - Un capellán no suele ser bebedor de cerveza.
 - Un bebedor de cerveza suele tener sobrepeso.
 - Por lo general, un infante de la marina está en buena condición física.

Si se representa la información en una red de herencia se puede concluir que:

 - a) Existen dos razonamientos cancelables.
 - b) Juan tiene sobrepeso.
 - c) No se puede asegurar que Juan sea capellán.
 - d) No se puede asegurar que Juan esté en buena condición física.
3. El uso de Hadoop y MapReduce en el contexto de la RI tiene como objetivo:
 - a) Facilitar el análisis en tiempo real de datos de redes sociales.
 - b) Automatizar el mantenimiento de sistemas de bases de datos.
 - c) Mejorar la eficiencia energética en centros de datos.
 - d) Procesar y analizar grandes conjuntos de datos para la RI.
4. La transición de la Web 1.0 a la Web 2.0 se caracterizó principalmente por:
 - a) La reducción en el uso de HTML y CSS en el desarrollo de sitios web.
 - b) La disminución de la importancia de los motores de búsqueda en la navegación web.
 - c) El cambio de páginas web estáticas a dinámicas, permitiendo la interacción del usuario y la generación de contenido.
 - d) El aumento en la velocidad de conexión a internet, que permitió una mejor calidad de las páginas web.
5. La integración de Hadoop y MapReduce en la RI trae como ventaja:
 - a) La posibilidad de la extracción de información relevante y la generación de resultados significativos de grandes conjuntos de datos.
 - b) La reducción de los costos operativos a cero.
 - c) La eliminación de la necesidad de sistemas de bases de datos.
 - d) La garantía de la privacidad absoluta de los datos procesados.
6. La Web 2.0 se caracteriza por:
 - a) Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
 - b) Las plataformas ofrecen experiencias a la medida, permitiendo a los usuarios personalizar sus perfiles, recibir recomendaciones ajustadas al contenido y participar en filtrado colaborativo.
 - c) La existencia de sitios web dinámicos e interactivos que permiten a los usuarios participar, comentar e interactuar tanto con los creadores de contenido como con otros usuarios.

- d) El uso de distintas tecnologías para crear experiencias web más interactivas y con mayor capacidad de respuesta.
7. ¿Qué algoritmos permiten obtener información de una red?
- Índices de centralidad.
 - Detección de comunidades.
 - Hypertext Induced Topic Selection (HITS).
 - PageRank.
8. En el contexto del procesamiento de grandes conjuntos de datos, se puede asegurar que:
- MapReduce no es adecuado para el procesamiento de datos no estructurados.
 - MapReduce solo puede manejar pequeñas cantidades de datos y no escala bien a grandes conjuntos de datos.
 - MapReduce es un enfoque para procesar datos de forma secuencial en un solo servidor para evitar problemas de concurrencia.
 - MapReduce divide una tarea en múltiples pasos de map y reduce que se ejecutan de forma secuencial en diferentes servidores para mejorar el rendimiento y la escalabilidad.
9. Existen varias estrategias consideradas como no recomendables para mejorar el SEO de un sitio web. Dentro de estas estrategias negativas, se encuentran:
- Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - Evitar el uso de las meta etiquetas y descripciones del sitio web para reflejar el contenido de la página.
 - Construir enlaces naturales de sitios web con autoridad y relevancia temática.
 - Incluir una cantidad excesiva de palabras clave irrelevantes para intentar manipular los rankings de búsqueda.
10. En el diseño de un videojuego de roles (RPG) en el que los personajes tienen atributos como salud, fuerza y velocidad, ¿cuál de las siguientes opciones representa mejor una implementación de la representación del conocimiento orientado a objetos?
- Cada personaje se representa como un objeto con propiedades como salud, fuerza y velocidad, y métodos para modificar y consultar estos valores.
 - Cada personaje se representa como una función que calcula sus atributos en función de su nivel y experiencia.
 - Cada personaje se representa como una matriz de números que almacena sus valores de atributos.
 - Cada personaje se representa como una lista de cadenas de texto que describen sus características físicas y habilidades.
11. En el contexto de la RI en redes, se puede afirmar que:
- La detección de comunidades ayuda a identificar grupos de nodos altamente conectados entre sí, lo que puede ser útil para comprender la estructura y el contenido de la red.
 - La detección de comunidades en una red siempre produce resultados objetivos y consistentes independientemente del algoritmo utilizado.
 - La detección de comunidades solo se aplica a redes pequeñas y simples, no a redes grandes y complejas.
 - La detección de comunidades no es relevante para la RI en redes, ya que se centra únicamente en la estructura de la red sin considerar el contenido.
12. Si se tiene el conjunto de páginas interconectadas $\{A \rightarrow B, C; B \rightarrow C; C \rightarrow A\}$, entonces la página con valor más alto de PageRank es:
- B.
 - C.
 - A y B.
 - A.
13. ¿Cuál de las siguientes opciones describe mejor la diferencia clave entre Web Crawling y Web Scraping?
- Web Scraping implica el análisis de la estructura y el contenido de las páginas web para extraer datos, mientras que Web Crawling se refiere a la descarga y almacenamiento de páginas web completas.
 - Web Scraping es más eficaz para rastrear e indexar contenido web para motores de búsqueda, mientras que Web Crawling se utiliza principalmente para la extracción de datos en proyectos de investigación.
 - Web Crawling se centra en la exploración y recopilación de enlaces de múltiples sitios web, mientras que Web Scraping se enfoca en la extracción específica de datos de páginas web individuales.
 - Web Crawling se realiza utilizando herramientas de automatización como Selenium WebDriver, mientras que Web Scraping se lleva a cabo mediante el análisis de HTML y CSS.

14. Una plataforma de comercio electrónico desea mejorar la experiencia del usuario al permitir una navegación más personalizada y contextualizada. Actualmente los usuarios tienen dificultades para encontrar productos relevantes debido a la gran cantidad de opciones disponibles. La empresa está interesada en implementar características de la Web 2.5 y la Web Semántica para abordar este problema. ¿Qué características podrían ayudar para ofrecer una navegación más personalizada y contextualizada?
- a) La implementación de ontologías y metadatos para enriquecer la descripción de productos y mejorar la precisión de las recomendaciones.
 - b) La optimización de la velocidad de carga del sitio web para mejorar la experiencia del usuario y reducir el abandono del carrito de compra.
15. En un SRI la indexación:
- a) Mejora la experiencia del usuario.
 - b) Consiste en asociar un identificador único a cada dato almacenado en el sistema.
 - c) Facilita la RI relevante.
 - d) Permite la organización y la categorización de la información.
- c) La incorporación de los productos en tendencia en el mercado.
 - d) La integración de redes sociales para permitir la recomendación de productos basada en las preferencias de amigos y contactos.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Un investigador necesita recopilar datos de múltiples sitios web para un estudio académico, pero se enfrenta a varios desafíos al realizar el proceso de Web Scraping de manera ética y legal. ¿Cuál de las siguientes opciones describe mejor uno de los desafíos asociados al proceso de Web Scraping?
 - a) La dificultad para encontrar herramientas de Web Scraping gratuitas y fiables.
 - b) La disponibilidad limitada de datos en línea que se pueden extraer utilizando técnicas de Web Scraping.
 - c) La necesidad de estar montado sobre un crawler que cumpla con todas las políticas.
 - d) La necesidad de comprender la estructura del sitio web y su código HTML para extraer los datos correctamente.
2. En una red de transporte donde cada nodo es una parada de autobús y las aristas representan si existe un carro que pasa por ambos sitios, ¿qué puede mejorar el sistema de transporte?
 - a) Utilizar la centralidad de intermediación para identificar las paradas de transferencia clave y establecer nuevas conexiones entre diferentes líneas de transporte público.
 - b) Utilizar el grafo inducido de los nodos con mayor valor en la centralidad de grado para aplicar la centralidad de intermediación con el propósito de reforzar las paradas con mayor tráfico.
 - c) No implementar sistemas de información en tiempo real para los usuarios, ya que pueden aumentar la carga de los trabajadores.
 - d) No considerar la centralidad de intermediación de las estaciones de transporte público al planificar rutas y horarios, ya que no tiene impacto en las conexiones entre las paradas.
3. El algoritmo de PageRank converge si:
 - a) Se define un factor de normalización en la fórmula de la función.
 - b) La norma de la diferencia entre los vectores es menor a un umbral predefinido.
 - c) Finaliza la ejecución cuando el número de iteraciones excede un máximo de iteraciones previamente definido.
 - d) El algoritmo no se implementa de forma iterativa.
4. La diferencia entre la indexación por tokens y la indexación por conceptos puede definirse como:
 - a) La indexación por tokens normaliza los datos reduciéndolos a su forma básica, mientras que la indexación por conceptos utiliza un método de ordenación para organizar los términos característicos de los datos.
 - b) La indexación por tokens asigna un valor numérico a cada término de los datos, mientras que la indexación por conceptos utiliza algoritmos de encriptación para proteger la privacidad de los datos.
 - c) La indexación por tokens asigna pesos a los términos basados en su importancia relativa, mientras que la indexación por conceptos utiliza un sistema de etiquetado para asociar términos con características generales.
 - d) La indexación por tokens divide los datos en términos individuales, mientras que la indexación por conceptos agrupa los datos en categorías definidas.
5. ¿Qué algoritmos permiten obtener información de una red?
 - a) Índices de centralidad.
 - b) Detección de comunidades.
 - c) Hypertext Induced Topic Selection (HITS).
 - d) PageRank.
6. Analizar una red permite:
 - a) Evaluar la calidad del contenido de un sitio web.
 - b) Obtener predicciones exactas de eventos futuros en mercados financieros.
 - c) Detectar posibles tendencias antes de que se conviertan en tendencia.
 - d) Encontrar nodos “sensibles” o críticos para la red.
7. El algoritmo Hypertext Induced Topic Selection (HITS) intenta buscar nodos especiales. Estos son conocidos como:
 - a) Sensibles.
 - b) Centrales.

- c) Hubs.
d) Autoridades.
8. El concepto de *rank sink* en el algoritmo de PageRank representa páginas web:
- Con una baja calidad de contenido y una cantidad insuficiente de enlaces salientes, lo que las hace menos relevantes en los resultados de búsqueda.
 - Con distribuciones uniformes de la puntuación de PageRank.
 - Con un alto número de enlaces salientes que no reciben enlaces entrantes, lo que puede afectar negativamente su puntuación de PageRank.
 - Que tienen una puntuación de PageRank más alta que otras debido a la manipulación de enlaces entrantes y salientes.
9. La afirmación que mejor describe la política de amabilidad en los Web Crawlers es:
- Los Web Crawlers son libres de recopilar datos de cualquier sitio web sin restricciones, independientemente de la cantidad de tráfico que generen.
 - La política de amabilidad establece pautas y reglas sobre cómo los crawlers deben interactuar con los sitios web para minimizar la carga del servidor y respetar las directivas de los administradores del sitio.
 - La política de amabilidad de los Web Crawlers dicta que los crawlers deben priorizar ciertos tipos de contenido sobre otros, ignorando completamente ciertas páginas web.
 - Los crawlers se diseñan para acceder a sitios web sin restricciones y extraer datos de manera agresiva para su indexación.
10. La centralidad de intermediación de un nodo indica:
- La frecuencia con la que un nodo actúa como puente en el camino más corto entre otros dos nodos.
 - El número total de conexiones entrantes y salientes.
 - La resistencia del nodo a fallos.
 - La cantidad de nodos vecinos directos.
11. Un sistema cuenta con la siguiente información:
- Juan pertenece a la marina.
 - Juan es capellán (sacerdote encargado de una tarea específica fuera de la parroquia).
 - Los infantes de la marina suelen ser bebedores de cerveza.
 - Un capellán no suele ser bebedor de cerveza.
 - Un bebedor de cerveza suele tener sobrepeso.
 - Por lo general, un infante de la marina está en buena condición física.
- Si se representa la información en una red de herencia se puede concluir que:
- Existen dos razonamientos cancelables.
 - Juan tiene sobrepeso.
 - No se puede asegurar que Juan sea capellán.
 - No se puede asegurar que Juan esté en buena condición física.
12. El componente responsable de la gestión de recursos y planificación de tareas en Hadoop es:
- MapReduce.
 - HDFS.
 - Hadoop Common.
 - YARN.
13. El uso de Hadoop y MapReduce en el contexto de la RI tiene como objetivo:
- Facilitar el análisis en tiempo real de datos de redes sociales.
 - Automatizar el mantenimiento de sistemas de bases de datos.
 - Mejorar la eficiencia energética en centros de datos.
 - Procesar y analizar grandes conjuntos de datos para la RI.
14. Al realizar la optimización de contenido para SEO debe considerarse:
- Utilizar etiquetas de título y meta descripciones únicas y relevantes para cada página.
 - Seleccionar las palabras clave al azar.
 - Crear contenido valioso y original que satisfaga las necesidades de los usuarios.
 - Incluir palabras clave de manera excesiva en el contenido para mejorar el posicionamiento.
15. Si se tiene el conjunto de páginas interconectadas $\{A \rightarrow B, C; B \rightarrow C; C \rightarrow A\}$, entonces la página con valor más alto de PageRank es:
- B.
 - C.
 - A y B.
 - A.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Dentro del análisis de redes, la centralidad de grado mide la importancia de un nodo basándose en:
 - a) La cantidad de vecinos del nodo.
 - b) El número de aristas que posee el nodo.
 - c) La cantidad de veces que aparece el nodo en el camino mínimo entre cualquier par de nodos.
 - d) El grado del nodo.
2. Si se tiene el conjunto de páginas interconectadas $\{A \rightarrow B, C; B \rightarrow C; C \rightarrow A\}$, entonces la página con valor más alto de PageRank es:
 - a) B.
 - b) C.
 - c) A y B.
 - d) A.
3. Se tiene una red de ingredientes donde cada uno representa un nodo y las aristas simbolizan que los ingredientes forman parte de una misma receta. Se busca mejorar la experiencia culinaria mediante la elaboración de combinaciones de ingredientes más interesantes y creativas, para lo cual se debe:
 - a) No tener en cuenta la centralidad de grado de los ingredientes para no darle mayor importancia a los ingredientes más comunes.
 - b) No considerar la centralidad de grado de los ingredientes, ya que todas las combinaciones de ingredientes son igualmente válidas.
 - c) Utilizar la centralidad de grado para identificar los ingredientes más populares en la red y crear combinaciones que incluyan una variedad de ingredientes menos comunes.
 - d) Utilizar la centralidad de grado para identificar los ingredientes menos conectados en la red y tomarlos en cuenta para su inclusión en las combinaciones.
4. Dentro del ecosistema de Hadoop, el HDFS se caracteriza por:
 - a) La capacidad ilimitada de almacenamiento.
 - b) La tolerancia a fallos mediante la replicación de datos.
 - c) El almacenamiento exclusivo para archivos de texto.
 - d) El modelo de acceso y de escritura de datos en tiempo real.
5. En una empresa comercial, se tiene una red donde los nodos corresponden a los empleados y las aristas representan la frecuencia con la que colaboran en las ventas. Se busca mejorar la colaboración entre los empleados para aumentar las ventas totales, por lo que la directiva debe:
 - a) No tener en cuenta la centralidad de cercanía de los empleados al diseñar estrategias de colaboración, ya que esto podría introducir complicaciones adicionales en el proceso de trabajo.
 - b) Utilizar la centralidad de cercanía para identificar los empleados menos cercanos a otros en la red y asignarles tareas individuales para evitar posibles conflictos y desacuerdos en el proceso de colaboración.
 - c) No considerar la centralidad de cercanía de los empleados al diseñar estrategias de colaboración, ya que todos los empleados tienen el mismo acceso a la información y los recursos.
 - d) Utilizar la centralidad de cercanía para identificar los empleados más cercanos a otros en la red y promover la colaboración entre ellos, facilitando así la comunicación y el intercambio de conocimientos para mejorar las ventas.
6. El conjunto frontera de URLs en un Web Crawler:
 - a) Delimita los sitios web que visitará en futuras iteraciones del proceso.
 - b) Indica las secciones que no pueden ser visitadas de cada sitio web.
 - c) Almacena hipervínculos que pertenecen al mismo dominio del conjunto semilla de URLs con que inició el crawler.
 - d) Es similar a un conjunto de URLs que esperan ser visitadas.
7. Un SRI es capaz de:
 - a) No necesitar de ningún almacenamiento externo para alojar los índices de los datos.
 - b) Reducir el tiempo de indexación de los datos si utiliza servidores distribuidos dentro de la red para que cada uno ejecute la indexación del mismo conjunto de datos.

- c) Generar índices invertidos de manera óptima sin considerar el contexto.
 - d) Crear los índices asociados a los datos sin tener que analizar cada dato.
8. La centralidad de intermediación de un nodo indica:
- a) La frecuencia con la que un nodo actúa como puente en el camino más corto entre otros dos nodos.
 - b) El número total de conexiones entrantes y salientes.
 - c) La resistencia del nodo a fallos.
 - d) La cantidad de nodos vecinos directos.
9. En el diseño de un videojuego de roles (RPG) en el que los personajes tienen atributos como salud, fuerza y velocidad, ¿cuál de las siguientes opciones representa mejor una implementación de la representación del conocimiento orientado a objetos?
- a) Cada personaje se representa como un objeto con propiedades como salud, fuerza y velocidad, y métodos para modificar y consultar estos valores.
 - b) Cada personaje se representa como una función que calcula sus atributos en función de su nivel y experiencia.
 - c) Cada personaje se representa como una matriz de números que almacena sus valores de atributos.
 - d) Cada personaje se representa como una lista de cadenas de texto que describen sus características físicas y habilidades.
10. La afirmación que mejor refleja el principio subyacente de PageRank, considerando su importancia en la clasificación de los sitios web, es:
- a) El algoritmo de PageRank considera tanto la cantidad como la calidad de los enlaces entrantes, asignando mayor valor a los enlaces provenientes de sitios web considerados como “importantes”.
 - b) PageRank valora más la cantidad de enlaces entrantes a una página web, independientemente de la calidad o relevancia de estos enlaces.
 - c) La efectividad de PageRank se basa exclusivamente en el análisis de las palabras clave contenidas en los enlaces entrantes, sin considerar la estructura de enlace de la Web.
 - d) PageRank opera bajo el supuesto de que los enlaces entrantes y salientes tienen el mismo impacto en la valoración de la relevancia de una página web.
11. La transición de la Web 1.0 a la Web 2.0 se caracterizó principalmente por:
- a) La reducción en el uso de HTML y CSS en el desarrollo de sitios web.
 - b) La disminución de la importancia de los motores de búsqueda en la navegación web.
 - c) El cambio de páginas web estáticas a dinámicas, permitiendo la interacción del usuario y la generación de contenido.
 - d) El aumento en la velocidad de conexión a internet, que permitió una mejor calidad de las páginas web.
12. En el contexto del procesamiento de grandes conjuntos de datos, se puede asegurar que:
- a) MapReduce no es adecuado para el procesamiento de datos no estructurados.
 - b) MapReduce solo puede manejar pequeñas cantidades de datos y no escala bien a grandes conjuntos de datos.
 - c) MapReduce es un enfoque para procesar datos de forma secuencial en un solo servidor para evitar problemas de concurrencia.
 - d) MapReduce divide una tarea en múltiples pasos de map y reduce que se ejecutan de forma secuencial en diferentes servidores para mejorar el rendimiento y la escalabilidad.
13. Una empresa de comercio electrónico necesita mejorar su motor de búsqueda para proporcionar resultados más relevantes a sus usuarios. Actualmente, los resultados de la búsqueda no son precisos y los usuarios a menudo encuentran dificultades para encontrar productos específicos. La empresa está considerando implementar una indexación por conceptos para mejorar la relevancia de los resultados de búsqueda. ¿Qué beneficios podría provocar este cambio?
- a) Permite adaptarse fácilmente a cambios en el vocabulario y la terminología utilizada en las descripciones de los productos.
 - b) Facilita la visualización de productos al agruparlos por categorías o características comunes.
 - c) Mejora la precisión en la búsqueda de productos relacionados, incluso cuando no coinciden exactamente con los términos de búsqueda del usuario.
 - d) Ayuda a identificar automáticamente relaciones entre productos, lo que puede mejorar las recomendaciones personalizadas a los usuarios.
14. En la RI en el contexto de Big Data, se puede asegurar que:
- a) El almacenamiento distribuido es una técnica obsoleta.

- b) El almacenamiento distribuido centraliza todos los datos en un único servidor para facilitar su gestión y mantenimiento.
- c) El almacenamiento distribuido reparte los datos en múltiples servidores para mejorar la disponibilidad y la redundancia del sistema.
- d) El almacenamiento distribuido ofrece ventajas significativas en términos de escalabilidad y rendimiento en comparación con el almacenamiento centralizado.
15. Para la RI, el análisis de las redes puede:
- a) Ayudar a identificar grupos de interés.
- b) Indicar la importancia de una entidad en la transmisión de la información.
- c) Revelar patrones de influencia dentro de una comunidad.
- d) Ayudar a comprender la conectividad y la accesibilidad entre las entidades.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Referente al Web Crawler puede afirmarse que:
 - a) Los hipervínculos encontrados en cada sitio web que no pertenecen al dominio donde fueron encontrados se desechan, puesto que no expande el conjunto de URLs sin visitar.
 - b) Las páginas visitadas no se procesan nunca más.
 - c) No necesita de un conjunto inicial de URLs para recorrer la Web.
 - d) No tiene como objetivo indexar y recopilar información de diferentes sitios web.
2. La web actual se enfrenta a problemas como:
 - a) Presencia de una alta calidad en los datos.
 - b) Heterogeneidad en los datos.
 - c) Presencia de datos volátiles y distribuidos.
 - d) Presencia de grandes volúmenes de datos estructurados.
3. ¿Cuál de las siguientes opciones describe mejor la diferencia clave entre Web Crawling y Web Scraping?
 - a) Web Scraping implica el análisis de la estructura y el contenido de las páginas web para extraer datos, mientras que Web Crawling se refiere a la descarga y almacenamiento de páginas web completas.
 - b) Web Scraping es más eficaz para rastrear e indexar contenido web para motores de búsqueda, mientras que Web Crawling se utiliza principalmente para la extracción de datos en proyectos de investigación.
 - c) Web Crawling se centra en la exploración y recopilación de enlaces de múltiples sitios web, mientras que Web Scraping se enfoca en la extracción específica de datos de páginas web individuales.
 - d) Web Crawling se realiza utilizando herramientas de automatización como Selenium WebDriver, mientras que Web Scraping se lleva a cabo mediante el análisis de HTML y CSS.
4. Sobre el algoritmo de PageRank, visto en clase, se puede afirmar que:
 - a) Evalúa la importancia de un sitio web en función de la calidad y cantidad de enlaces entrantes que recibe de otros sitios web.
 - b) Asigna una puntuación baja a los sitios web que contienen muchos enlaces salientes, ya que indica una falta de relevancia.
 - c) Asigna una puntuación alta a los sitios web que tienen un gran número de enlaces entrantes sin tener en cuenta la calidad de esos enlaces.
 - d) Solo tiene en cuenta el contenido en un sitio web para determinar su relevancia en los resultados de búsqueda.
5. La centralidad de intermediación de un nodo indica:
 - a) La frecuencia con la que un nodo actúa como puente en el camino más corto entre otros dos nodos.
 - b) El número total de conexiones entrantes y salientes.
 - c) La resistencia del nodo a fallos.
 - d) La cantidad de nodos vecinos directos.
6. Una buena práctica de SEO para mejorar el posicionamiento de un sitio web en los motores de búsqueda es:
 - a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Copiar contenido directamente de otros sitios web populares para aumentar la cantidad de páginas indexadas.
 - c) Llenar el contenido con palabras clave irrelevantes para aumentar la densidad de palabras clave.
 - d) Obtener enlaces de otros sitios web relevantes y de calidad que apunten al sitio.
7. En el contexto de la representación del conocimiento basada en herencia, ¿qué caracteriza a la herencia cancelable?
 - a) Las propiedades heredadas siempre se mantienen y no pueden anularse.
 - b) La herencia es el resultado del razonamiento no transitivo.

- c) Las conclusiones no están determinadas y dependen del nodo de interés.
 - d) No existe ambigüedad en las conclusiones obtenidas.
8. El componente responsable de la gestión de recursos y planificación de tareas en Hadoop es:
- a) MapReduce.
 - b) HDFS.
 - c) Hadoop Common.
 - d) YARN.
9. En una biblioteca digital se necesita implementar un sistema que permita a los usuarios encontrar libros y artículos científicos de forma rápida y precisa. Los documentos están en diversos formatos, incluyendo PDF, EPUB y HTML. Se requiere seleccionar un algoritmo de indexación adecuado para el sistema, por lo que el programador designado para la implementación debe considerar:
- a) La capacidad para manejar documentos en diferentes formatos.
 - b) La velocidad de indexación y recuperación de datos.
 - c) La complejidad del algoritmo en términos de implementación y mantenimiento.
 - d) La capacidad del algoritmo para procesar imágenes incrustadas o referenciadas en los ficheros.
10. La política de ordenación de URLs en los Web Crawlers tiene como aspecto fundamental:
- a) Limitar el acceso de los crawlers a ciertas secciones de un sitio web, evitando el rastreo de URLs consideradas menos importantes o sensibles.
 - b) Establecer la prioridad de rastreo de las URLs, determinando el orden en que los crawlers visitan y procesan cada página web.
 - c) Definir la estructura de la URL de destino, asegurando que estén ordenadas alfabéticamente para facilitar la navegación y la indexación.
 - d) Determinar la forma en que los crawlers asignan un valor de relevancia a cada URL para clasificarlas en el índice de búsqueda.
11. Analizar una red permite:
- a) Evaluar la calidad del contenido de un sitio web.
 - b) Obtener predicciones exactas de eventos futuros en mercados financieros.
 - c) Detectar posibles tendencias antes de que se conviertan en tendencia.
 - d) Encontrar nodos “sensibles” o críticos para la red.
12. En un sistema donde el conocimiento está orientado a objetos se puede asegurar que:
- a) Los marcos y las bandas son las estructuras utilizadas para la representación del modelo.
 - b) Las funciones de agregación dificultan poder establecer relaciones entre los objetos.
 - c) Se enfatiza la atención a la información de la cual se extrajo el conocimiento.
 - d) No es posible definir especializaciones de los objetos de la vida real dentro del sistema.
13. Sobre WordNet y su aplicación en el procesamiento del lenguaje natural, se puede afirmar que:
- a) Aunque WordNet es una herramienta valiosa en el NLP, su estructura no incluye información sobre la frecuencia de uso de las palabras en el lenguaje natural.
 - b) El diseño de WordNet facilita su integración en aplicaciones multilingües de NLP, aunque su desarrollo original se centró en el inglés.
 - c) Los synsets facilitan la identificación de relaciones semánticas entre palabras, como la hiperonimia y la meronimia, enriqueciendo tareas de NLP.
 - d) WordNet proporciona una base para la desambiguación semántica al agrupar palabras con significados similares en synsets.
14. En la RI en el contexto de Big Data, se puede asegurar que:
- a) El almacenamiento distribuido es una técnica obsoleta.
 - b) El almacenamiento distribuido centraliza todos los datos en un único servidor para facilitar su gestión y mantenimiento.
 - c) El almacenamiento distribuido reparte los datos en múltiples servidores para mejorar la disponibilidad y la redundancia del sistema.
 - d) El almacenamiento distribuido ofrece ventajas significativas en términos de escalabilidad y rendimiento en comparación con el almacenamiento centralizado.
15. Dentro del análisis de redes, la centralidad de grado mide la importancia de un nodo basándose en:
- a) La cantidad de vecinos del nodo.
 - b) El número de aristas que posee el nodo.
 - c) La cantidad de veces que aparece el nodo en el camino mínimo entre cualquier par de nodos.
 - d) El grado del nodo.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. ¿Qué es un 'Uniform Resource Locator (URL) Frontier' en el contexto de Web Crawling?
 - a) Una base de datos que almacena URLs únicas identificadas como recursos en la web.
 - b) Una técnica para filtrar URLs irrelevantes y mejorar la eficiencia del crawling.
 - c) Un protocolo que define cómo se deben formatear las URLs para el crawling.
 - d) Una lista prioritaria de URLs que aún no han sido visitadas por el crawler.
2. Un SRI es capaz de:
 - a) No necesitar de ningún almacenamiento externo para alojar los índices de los datos.
 - b) Reducir el tiempo de indexación de los datos si utiliza servidores distribuidos dentro de la red para que cada uno ejecute la indexación del mismo conjunto de datos.
 - c) Generar índices invertidos de manera óptima sin considerar el contexto.
 - d) Crear los índices asociados a los datos sin tener que analizar cada dato.
3. Un investigador necesita recopilar datos de múltiples sitios web para un estudio académico, pero se enfrenta a varios desafíos al realizar el proceso de Web Scraping de manera ética y legal. ¿Cuál de las siguientes opciones describe mejor uno de los desafíos asociados al proceso de Web Scraping?
 - a) La dificultad para encontrar herramientas de Web Scraping gratuitas y fiables.
 - b) La disponibilidad limitada de datos en línea que se pueden extraer utilizando técnicas de Web Scraping.
 - c) La necesidad de estar montado sobre un crawler que cumpla con todas las políticas.
 - d) La necesidad de comprender la estructura del sitio web y su código HTML para extraer los datos correctamente.
4. Para la RI, el análisis de las redes puede:
 - a) Ayudar a identificar grupos de interés.
 - b) Indicar la importancia de una entidad en la transmisión de la información.
 - c) Revelar patrones de influencia dentro de una comunidad.
 - d) Ayudar a comprender la conectividad y la accesibilidad entre las entidades.
5. La Web 1.0 se caracteriza por:
 - a) Los propietarios de los sitios web proporcionan contenido de forma periódica.
 - b) Uso de comunidades virtuales para popularizar los sitios web de noticias.
 - c) Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
 - d) Los sitios web se centran en brindar información en lugar de facilitar la colaboración o participación de los usuarios.
6. Dentro del análisis de redes, la centralidad de grado mide la importancia de un nodo basándose en:
 - a) La cantidad de vecinos del nodo.
 - b) El número de aristas que posee el nodo.
 - c) La cantidad de veces que aparece el nodo en el camino mínimo entre cualquier par de nodos.
 - d) El grado del nodo.
7. Considerando las prácticas éticas y legales en el Web Scraping, se puede asegurar que:
 - a) Web Scraping sobre datos personales sin consentimiento es generalmente aceptado si los datos se utilizan con fines de investigación.
 - b) Es importante revisar y respetar los términos de servicio del sitio web, así como las leyes aplicables de protección de datos y derechos de autor, antes de realizar Web Scraping.
 - c) Web Scraping requiere considerar las políticas de `robots.txt` del sitio web objetivo.
 - d) La extracción de datos mediante Web Scraping siempre es legal, independientemente de las leyes locales sobre derechos de autor y privacidad de datos definidos en los sitios web.
8. El algoritmo de PageRank puede describirse como un procedimiento utilizado para:
 - a) Determinar la velocidad de carga de una página web en un navegador.

- b) Calcular la relevancia de una página web en función de la cantidad y calidad de los enlaces que apuntan hacia ella.
 - c) Clasificar las páginas web en función de su edad y autoridad.
 - d) Prevenir el spam y el contenido no deseado en las páginas web.
9. El uso de Hadoop y MapReduce en el contexto de la RI tiene como objetivo:
- a) Facilitar el análisis en tiempo real de datos de redes sociales.
 - b) Automatizar el mantenimiento de sistemas de bases de datos.
 - c) Mejorar la eficiencia energética en centros de datos.
 - d) Procesar y analizar grandes conjuntos de datos para la RI.
10. En un sistema de control de tráfico urbano basado en reglas, ¿cuál de las siguientes reglas sería más efectiva para manejar situaciones de congestión en una intersección?
- a) Si hay pocos vehículos en la intersección, reducir el tiempo de los semáforos en verde.
 - b) Si hay muchos vehículos en la intersección, aumentar el tiempo de los semáforos en verde.
 - c) Si hay un vehículo de emergencia en la intersección, detener todos los demás vehículos.
 - d) Si hay muchos peatones cruzando la intersección, reducir el tiempo de los semáforos en rojo.
11. El concepto de *rank sink* en el algoritmo de PageRank representa páginas web:
- a) Con una baja calidad de contenido y una cantidad insuficiente de enlaces salientes, lo que las hace menos relevantes en los resultados de búsqueda.
 - b) Con distribuciones uniformes de la puntuación de PageRank.
 - c) Con un alto número de enlaces salientes que no reciben enlaces entrantes, lo que puede afectar negativamente su puntuación de PageRank.
 - d) Que tienen una puntuación de PageRank más alta que otras debido a la manipulación de enlaces entrantes y salientes.
12. Una plataforma de comercio electrónico desea mejorar la experiencia del usuario al permitir una navegación más personalizada y contextualizada. Actualmente los usuarios tienen dificultades para encontrar productos relevantes debido a la gran cantidad de opciones disponibles. La empresa está interesada en implementar características de la Web 2.5 y la Web Semántica para abordar este problema. ¿Qué características podrían ayudar para ofrecer una navegación más personalizada y contextualizada?
- a) La implementación de ontologías y metadatos para enriquecer la descripción de productos y mejorar la precisión de las recomendaciones.
 - b) La optimización de la velocidad de carga del sitio web para mejorar la experiencia del usuario y reducir el abandono del carrito de compra.
 - c) La incorporación de los productos en tendencia en el mercado.
 - d) La integración de redes sociales para permitir la recomendación de productos basada en las preferencias de amigos y contactos.
13. En la RI en el contexto de Big Data, se puede asegurar que:
- a) La indexación distribuida es una técnica obsoleta.
 - b) La indexación distribuida almacena todos los datos en un solo servidor para facilitar su acceso y búsqueda.
 - c) La indexación distribuida divide los datos en múltiples fragmentos que se almacenan en varios nodos para permitir búsquedas paralelas y mejorar la escalabilidad.
 - d) La indexación distribuida no ofrece ventajas en términos de rendimiento y escalabilidad en comparación con la indexación centralizada.
14. El algoritmo de PageRank converge si:
- a) Se define un factor de normalización en la fórmula de la función.
 - b) La norma de la diferencia entre los vectores es menor a un umbral predefinido.
 - c) Finaliza la ejecución cuando el número de iteraciones excede un máximo de iteraciones previamente definido.
 - d) El algoritmo no se implementa de forma iterativa.
15. En el diseño de un videojuego de roles (RPG) en el que los personajes tienen atributos como salud, fuerza y velocidad, ¿cuál de las siguientes opciones representa mejor una implementación de la representación del conocimiento orientado a objetos?
- a) Cada personaje se representa como un objeto con propiedades como salud, fuerza y velocidad, y métodos para modificar y consultar estos valores.
 - b) Cada personaje se representa como una función que calcula sus atributos en función de su nivel y experiencia.

- c) Cada personaje se representa como una matriz de números que almacena sus valores de atributos.
- d) Cada personaje se representa como una lista de cadenas de texto que describen sus características físicas y habilidades.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. En una biblioteca digital se necesita implementar un sistema que permita a los usuarios encontrar libros y artículos científicos de forma rápida y precisa. Los documentos están en diversos formatos, incluyendo PDF, EPUB y HTML. Se requiere seleccionar un algoritmo de indexación adecuado para el sistema, por lo que el programador designado para la implementación debe considerar:
 - a) La capacidad para manejar documentos en diferentes formatos.
 - b) La velocidad de indexación y recuperación de datos.
 - c) La complejidad del algoritmo en términos de implementación y mantenimiento.
 - d) La capacidad del algoritmo para procesar imágenes incrustadas o referenciadas en los ficheros.
2. Si se tiene el conjunto de páginas interconectadas $\{A \rightarrow B, C; B \rightarrow C; C \rightarrow A\}$, entonces la página con valor más alto de PageRank es:
 - a) B.
 - b) C.
 - c) A y B.
 - d) A.
3. Considerando las dimensiones y desafíos inherentes a Big Data puede afirmarse, tomando en cuenta las características clave y las implicaciones para su procesamiento y análisis, que:
 - a) Big Data se caracteriza principalmente por su pequeño volumen y uniformidad, permitiendo un procesamiento eficiente con mínimas adaptaciones de las herramientas de análisis de datos tradicionales.
 - b) Big Data no desafía la capacidad de las herramientas tradicionales de procesamiento de datos para capturar, almacenar, gestionar y analizar efectivamente la información, dada la evolución constante de las capacidades computacionales y algoritmos de optimización.
 - c) Aunque Big Data puede incluir datos estructurados, su naturaleza se expande al incorporar grandes cantidades de datos no estructurados y semiestructurados, lo que exige el uso de tecnologías especializadas en almacenamiento y procesamiento como Hadoop y sistemas de bases de datos NoSQL.
- d) Además de su complejidad y diversidad, Big Data introduce desafíos significativos en términos de veracidad y variabilidad al requerir métodos avanzados de limpieza y validación de datos para asegurar la integridad del análisis.
4. Una empresa de comercio electrónico necesita mejorar su motor de búsqueda para proporcionar resultados más relevantes a sus usuarios. Actualmente, los resultados de la búsqueda no son precisos y los usuarios a menudo encuentran dificultades para encontrar productos específicos. La empresa está considerando implementar una indexación por conceptos para mejorar la relevancia de los resultados de búsqueda. ¿Qué beneficios podría provocar este cambio?
 - a) Permite adaptarse fácilmente a cambios en el vocabulario y la terminología utilizada en las descripciones de los productos.
 - b) Facilita la visualización de productos al agruparlos por categorías o características comunes.
 - c) Mejora la precisión en la búsqueda de productos relacionados, incluso cuando no coinciden exactamente con los términos de búsqueda del usuario.
 - d) Ayuda a identificar automáticamente relaciones entre productos, lo que puede mejorar las recomendaciones personalizadas a los usuarios.
5. La diferencia entre la indexación por tokens y la indexación por conceptos puede definirse como:
 - a) La indexación por tokens normaliza los datos reduciéndolos a su forma básica, mientras que la indexación por conceptos utiliza un método de ordenación para organizar los términos característicos de los datos.
 - b) La indexación por tokens asigna un valor numérico a cada término de los datos, mientras que la indexación por conceptos utiliza algoritmos de encriptación para proteger la privacidad de los datos.
 - c) La indexación por tokens asigna pesos a los términos basados en su importancia relativa,

mientras que la indexación por conceptos utiliza un sistema de etiquetado para asociar términos con características generales.

- d) La indexación por tokens divide los datos en términos individuales, mientras que la indexación por conceptos agrupa los datos en categorías definidas.
6. La transición de la Web 1.0 a la Web 2.0 se caracterizó principalmente por:
- a) La reducción en el uso de HTML y CSS en el desarrollo de sitios web.
 - b) La disminución de la importancia de los motores de búsqueda en la navegación web.
 - c) El cambio de páginas web estáticas a dinámicas, permitiendo la interacción del usuario y la generación de contenido.
 - d) El aumento en la velocidad de conexión a internet, que permitió una mejor calidad de las páginas web.
7. ¿Qué es un 'Uniform Resource Locator (URL) Frontier' en el contexto de Web Crawling?
- a) Una base de datos que almacena URLs únicas identificadas como recursos en la web.
 - b) Una técnica para filtrar URLs irrelevantes y mejorar la eficiencia del crawling.
 - c) Un protocolo que define cómo se deben formatear las URLs para el crawling.
 - d) Una lista prioritaria de URLs que aún no han sido visitadas por el crawler.
8. ¿Qué algoritmos permiten obtener información de una red?
- a) Índices de centralidad.
 - b) Detección de comunidades.
 - c) Hypertext Induced Topic Selection (HITS).
 - d) PageRank.
9. La afirmación que mejor refleja el principio subyacente de PageRank, considerando su importancia en la clasificación de los sitios web, es:
- a) El algoritmo de PageRank considera tanto la cantidad como la calidad de los enlaces entrantes, asignando mayor valor a los enlaces provenientes de sitios web considerados como "importantes".
 - b) PageRank valora más la cantidad de enlaces entrantes a una página web, independientemente de la calidad o relevancia de estos enlaces.
 - c) La efectividad de PageRank se basa exclusivamente en el análisis de las palabras clave contenidas en los enlaces entrantes, sin considerar la estructura de enlace de la Web.
 - d) PageRank opera bajo el supuesto de que los enlaces entrantes y salientes tienen el mismo impacto en la valoración de la relevancia de una página web.
10. Un sistema cuenta con la siguiente información:
- Juan pertenece a la marina.
 - Juan es capellán (sacerdote encargado de una tarea específica fuera de la parroquia).
 - Los infantes de la marina suelen ser bebedores de cerveza.
 - Un capellán no suele ser bebedor de cerveza.
 - Un bebedor de cerveza suele tener sobrepeso.
 - Por lo general, un infante de la marina está en buena condición física.
- Si se representa la información en una red de herencia se puede concluir que:
- a) Existen dos razonamientos cancelables.
 - b) Juan tiene sobrepeso.
 - c) No se puede asegurar que Juan sea capellán.
 - d) No se puede asegurar que Juan esté en buena condición física.
11. La integración de Hadoop y MapReduce en la RI trae como ventaja:
- a) La posibilidad de la extracción de información relevante y la generación de resultados significativos de grandes conjuntos de datos.
 - b) La reducción de los costos operativos a cero.
 - c) La eliminación de la necesidad de sistemas de bases de datos.
 - d) La garantía de la privacidad absoluta de los datos procesados.
12. En una red de transporte donde cada nodo es una parada de autobús y las aristas representan si existe un carro que pasa por ambos sitios, ¿qué puede mejorar el sistema de transporte?
- a) Utilizar la centralidad de intermediación para identificar las paradas de transferencia clave y establecer nuevas conexiones entre diferentes líneas de transporte público.
 - b) Utilizar el grafo inducido de los nodos con mayor valor en la centralidad de grado para aplicar la centralidad de intermediación con el propósito de reforzar las paradas con mayor tráfico.
 - c) No implementar sistemas de información en tiempo real para los usuarios, ya que pueden aumentar la carga de los trabajadores.
 - d) No considerar la centralidad de intermediación de las estaciones de transporte público al planificar rutas y horarios, ya que no tiene impacto en las conexiones entre las paradas.

13. El componente responsable de la gestión de recursos y planificación de tareas en Hadoop es:
- a) MapReduce.
 - b) HDFS.
 - c) Hadoop Common.
 - d) YARN.
14. Dentro del análisis de redes, la centralidad de grado mide la importancia de un nodo basándose en:
- a) La cantidad de vecinos del nodo.
 - b) El número de aristas que posee el nodo.
 - c) La cantidad de veces que aparece el nodo en el camino mínimo entre cualquier par de nodos.
 - d) El grado del nodo.
15. Para transformar el contenido no estructurado de las páginas web en datos estructurados, el Web Scraping:
- a) Convierte de forma automática imágenes a texto.
 - b) Interpreta el código JavaScript en tiempo real.
 - c) Extrae información basada en patrones de HTML/CSS.
 - d) Analiza los protocolos de red.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Para la RI, el análisis de las redes puede:
 - a) Ayudar a identificar grupos de interés.
 - b) Indicar la importancia de una entidad en la transmisión de la información.
 - c) Revelar patrones de influencia dentro de una comunidad.
 - d) Ayudar a comprender la conectividad y la accesibilidad entre las entidades.
2. Un sistema cuenta con la siguiente información:
 - Juan pertenece a la marina.
 - Juan es capellán (sacerdote encargado de una tarea específica fuera de la parroquia).
 - Los infantes de la marina suelen ser bebedores de cerveza.
 - Un capellán no suele ser bebedor de cerveza.
 - Un bebedor de cerveza suele tener sobrepeso.
 - Por lo general, un infante de la marina está en buena condición física.

Si se representa la información en una red de herencia se puede concluir que:

 - a) Existen dos razonamientos cancelables.
 - b) Juan tiene sobrepeso.
 - c) No se puede asegurar que Juan sea capellán.
 - d) No se puede asegurar que Juan esté en buena condición física.
3. La computación evolutiva:
 - a) Utiliza algoritmos para buscar soluciones óptimas en grandes espacios de búsqueda, lo que la hace adecuada para problemas de optimización en la RI.
 - b) Es útil solo para problemas de clasificación de documentos y no para otras tareas de RI en general.
 - c) No es aplicable en la RI debido a la complejidad de los algoritmos evolutivos.
 - d) Solo puede manejar conjuntos de datos pequeños y no es escalable a grandes volúmenes de datos.
4. Se tiene un grafo donde cada nodo es un personaje de cierto libro de cuentos y la existencia de las aristas está definida si dos personajes aparecen en el mismo cuento. Se puede asegurar que:
 - a) La centralidad de grado es útil para identificar los personajes con más conexiones dentro de la red, lo que puede indicar su importancia en el libro.
 - b) La centralidad de cercanía indica el grado de conexión de cada personaje con el resto de los personajes de los cuentos del libro.
 - c) La centralidad de intermediación identifica a los personajes que actúan como conectores entre personajes de cuentos distintos.
 - d) La centralidad de cercanía ofrece una relación entre la cantidad de vecinos de un nodo con respecto a la longitud máxima de un camino dentro del grafo partiendo del nodo en cuestión.
5. La premisa básica del algoritmo de PageRank para clasificar páginas web en los resultados de búsqueda es:
 - a) Los enlaces entrantes a una página web desde otras páginas contribuyen a su importancia y clasificación.
 - b) La antigüedad de la página web es el principal factor para determinar su clasificación.
 - c) La cantidad de visitas que recibe una página web determina su clasificación.
 - d) El contenido y la relevancia de las palabras clave en la página web determinan su posición.
6. El posicionamiento de un sitio web en los motores de búsqueda puede ser afectado por:
 - a) Mantener una estructura de URL clara y coherente.
 - b) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - c) Contenido sin valor en el sitio web.
 - d) Obtener enlaces de sitios web irrelevantes y de baja calidad.
7. Los Web Crawlers se enfrentan a desafíos constantes. Dentro de ellos se encuentran:

- a) La falta de acceso a la base de datos del servidor web para extraer información actualizada.
 - b) La modificación del código y la estructura del sitio web.
 - c) La incapacidad para interpretar correctamente el lenguaje de programación utilizado en el desarrollo de los sitios web.
 - d) La dificultad para generar el contenido dinámico en tiempo real.
8. Si se tiene el conjunto de páginas interconectadas $\{A \rightarrow B, C; B \rightarrow C; C \rightarrow A\}$, entonces la página con valor más alto de PageRank es:
- a) B.
 - b) C.
 - c) A y B.
 - d) A.
9. Existen varias estrategias consideradas como no recomendables para mejorar el SEO de un sitio web. Dentro de estas estrategias negativas, se encuentran:
- a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Evitar el uso de las meta etiquetas y descripciones del sitio web para reflejar el contenido de la página.
 - c) Construir enlaces naturales de sitios web con autoridad y relevancia temática.
 - d) Incluir una cantidad excesiva de palabras clave irrelevantes para intentar manipular los rankings de búsqueda.
10. Considerando las dimensiones y desafíos inherentes a Big Data puede afirmarse, tomando en cuenta las características clave y las implicaciones para su procesamiento y análisis, que:
- a) Big Data se caracteriza principalmente por su pequeño volumen y uniformidad, permitiendo un procesamiento eficiente con mínimas adaptaciones de las herramientas de análisis de datos tradicionales.
 - b) Big Data no desafía la capacidad de las herramientas tradicionales de procesamiento de datos para capturar, almacenar, gestionar y analizar efectivamente la información, dada la evolución constante de las capacidades computacionales y algoritmos de optimización.
 - c) Aunque Big Data puede incluir datos estructurados, su naturaleza se expande al incorporar grandes cantidades de datos no estructurados y semiestructurados, lo que exige el uso de tecnologías especializadas en almacenamiento y procesamiento como Hadoop y sistemas de bases de datos NoSQL.
 - d) Además de su complejidad y diversidad, Big Data introduce desafíos significativos en términos de veracidad y variabilidad al requerir métodos avanzados de limpieza y validación de datos para asegurar la integridad del análisis.
11. La integración de Hadoop y MapReduce en la RI trae como ventaja:
- a) La posibilidad de la extracción de información relevante y la generación de resultados significativos de grandes conjuntos de datos.
 - b) La reducción de los costos operativos a cero.
 - c) La eliminación de la necesidad de sistemas de bases de datos.
 - d) La garantía de la privacidad absoluta de los datos procesados.
12. En el contexto de la RI en redes sociales se puede afirmar que:
- a) El análisis de redes sociales se centra exclusivamente en la cantidad de seguidores que tiene un usuario en particular para determinar su influencia en la red.
 - b) El análisis de redes sociales no es útil para comprender la difusión de información en una red social específica, para ello se utiliza la medida de centralidad de vector propio.
 - c) El análisis de centralidad de intermediación se utiliza para identificar usuarios que son importantes en una red social debido a su posición como “puentes” entre diferentes grupos de usuarios.
 - d) El análisis de sentimientos se utiliza para determinar la popularidad de una publicación en redes sociales sin tener en cuenta la opinión de los usuarios.
13. Un investigador necesita recopilar datos de múltiples sitios web para un estudio académico, pero se enfrenta a varios desafíos al realizar el proceso de Web Scraping de manera ética y legal. ¿Cuál de las siguientes opciones describe mejor uno de los desafíos asociados al proceso de Web Scraping?
- a) La dificultad para encontrar herramientas de Web Scraping gratuitas y fiables.
 - b) La disponibilidad limitada de datos en línea que se pueden extraer utilizando técnicas de Web Scraping.
 - c) La necesidad de estar montado sobre un crawler que cumpla con todas las políticas.
 - d) La necesidad de comprender la estructura del sitio web y su código HTML para extraer los datos correctamente.

14. En una biblioteca digital se necesita implementar un sistema que permita a los usuarios encontrar libros y artículos científicos de forma rápida y precisa. Los documentos están en diversos formatos, incluyendo PDF, EPUB y HTML. Se requiere seleccionar un algoritmo de indexación adecuado para el sistema, por lo que el programador designado para la implementación debe considerar:
- a) La capacidad para manejar documentos en diferentes formatos.
 - b) La velocidad de indexación y recuperación de datos.
 - c) La complejidad del algoritmo en términos de implementación y mantenimiento.
 - d) La capacidad del algoritmo para procesar imágenes incrustadas o referenciadas en los ficheros.
15. La web actual se enfrenta a problemas como:
- a) Presencia de una alta calidad en los datos.
 - b) Heterogeneidad en los datos.
 - c) Presencia de datos volátiles y distribuidos.
 - d) Presencia de grandes volúmenes de datos estructurados.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Se puede afirmar que:
 - a) La cantidad de conexiones de un nodo siempre indica su influencia en la red.
 - b) Todas las relaciones en una red tienen la misma importancia para el análisis.
 - c) No es posible identificar subgrupos dentro de una red utilizando análisis de redes.
 - d) El tamaño de una red es siempre indicativo de su efectividad en la transmisión de información.
2. La diferencia entre la indexación por tokens y la indexación por conceptos puede definirse como:
 - a) La indexación por tokens normaliza los datos reduciéndolos a su forma básica, mientras que la indexación por conceptos utiliza un método de ordenación para organizar los términos característicos de los datos.
 - b) La indexación por tokens asigna un valor numérico a cada término de los datos, mientras que la indexación por conceptos utiliza algoritmos de encriptación para proteger la privacidad de los datos.
 - c) La indexación por tokens asigna pesos a los términos basados en su importancia relativa, mientras que la indexación por conceptos utiliza un sistema de etiquetado para asociar términos con características generales.
 - d) La indexación por tokens divide los datos en términos individuales, mientras que la indexación por conceptos agrupa los datos en categorías definidas.
3. En el modelo de representación del conocimiento basado en herencia se puede asegurar que:
 - a) La cantidad de padres que puede tener un nodo no es mayor que 4.
 - b) Solo se usa en entornos referentes a la biología.
 - c) El razonamiento deducido está respaldado por al menos un camino dentro del grafo.
 - d) Las conclusiones pueden ser canceladas si el grafo es ambiguo.
4. La web actual se enfrenta a problemas como:
 - a) Presencia de una alta calidad en los datos.
 - b) Heterogeneidad en los datos.
 - c) Presencia de datos volátiles y distribuidos.
 - d) Presencia de grandes volúmenes de datos estructurados.
5. En el contexto de la representación del conocimiento basada en herencia, ¿qué caracteriza a la herencia cancelable?
 - a) Las propiedades heredadas siempre se mantienen y no pueden anularse.
 - b) La herencia es el resultado del razonamiento no transitivo.
 - c) Las conclusiones no están determinadas y dependen del nodo de interés.
 - d) No existe ambigüedad en las conclusiones obtenidas.
6. El componente responsable de la gestión de recursos y planificación de tareas en Hadoop es:
 - a) MapReduce.
 - b) HDFS.
 - c) Hadoop Common.
 - d) YARN.
7. ¿Cuál de las siguientes opciones describe mejor la diferencia clave entre Web Crawling y Web Scraping?
 - a) Web Scraping implica el análisis de la estructura y el contenido de las páginas web para extraer datos, mientras que Web Crawling se refiere a la descarga y almacenamiento de páginas web completas.
 - b) Web Scraping es más eficaz para rastrear e indexar contenido web para motores de búsqueda, mientras que Web Crawling se utiliza principalmente para la extracción de datos en proyectos de investigación.
 - c) Web Crawling se centra en la exploración y recopilación de enlaces de múltiples sitios web, mientras que Web Scraping se enfoca en la extracción específica de datos de páginas web individuales.
 - d) Web Crawling se realiza utilizando herramientas de automatización como Selenium WebDriver, mientras que Web Scraping se lleva a cabo mediante el análisis de HTML y CSS.

8. El uso de Hadoop y MapReduce en el contexto de la RI tiene como objetivo:
 - a) Facilitar el análisis en tiempo real de datos de redes sociales.
 - b) Automatizar el mantenimiento de sistemas de bases de datos.
 - c) Mejorar la eficiencia energética en centros de datos.
 - d) Procesar y analizar grandes conjuntos de datos para la RI.
9. Si se tiene el conjunto de páginas interconectadas $\{A \rightarrow B, C; B \rightarrow C; C \rightarrow A\}$, entonces la página con valor más alto de PageRank es:
 - a) B.
 - b) C.
 - c) A y B.
 - d) A.
10. En un SRI la indexación:
 - a) Mejora la experiencia del usuario.
 - b) Consiste en asociar un identificador único a cada dato almacenado en el sistema.
 - c) Facilita la RI relevante.
 - d) Permite la organización y la categorización de la información.
11. En un sistema donde el conocimiento está orientado a objetos se puede asegurar que:
 - a) Los marcos y las bandas son las estructuras utilizadas para la representación del modelo.
 - b) Las funciones de agregación dificultan poder establecer relaciones entre los objetos.
 - c) Se enfatiza la atención a la información de la cual se extrajo el conocimiento.
 - d) No es posible definir especializaciones de los objetos de la vida real dentro del sistema.
12. ¿Qué estrategia utilizan los Web Crawlers para asegurar un rastreo eficiente y respetuoso de los recursos de los sitios web?
 - a) Extraer únicamente contenido multimedia para reducir la carga en los servidores web.
 - b) Visitar y rastrear todos los enlaces de una página web simultáneamente.
 - c) Ignorar completamente el archivo `robots.txt` de los sitios web.
 - d) Seguir las directrices del archivo `robots.txt` y aplicar un retraso entre las solicitudes.
13. Existen varias estrategias consideradas como no recomendables para mejorar el SEO de un sitio web. Dentro de estas estrategias negativas, se encuentran:
 - a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Evitar el uso de las meta etiquetas y descripciones del sitio web para reflejar el contenido de la página.
 - c) Construir enlaces naturales de sitios web con autoridad y relevancia temática.
 - d) Incluir una cantidad excesiva de palabras clave irrelevantes para intentar manipular los rankings de búsqueda.
14. Al “relajar” el concepto de clique en la detección de comunidades se intenta solucionar:
 - a) La uniformidad de los nodos en términos de grado.
 - b) La necesidad de datos externos para analizar la red.
 - c) La imposibilidad de conectar nodos distantes.
 - d) El solapamiento y la complejidad computacional.
15. Para transformar el contenido no estructurado de las páginas web en datos estructurados, el Web Scraping:
 - a) Convierte de forma automática imágenes a texto.
 - b) Interpreta el código JavaScript en tiempo real.
 - c) Extrae información basada en patrones de HTML/CSS.
 - d) Analiza los protocolos de red.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. La Web 2.0 se caracteriza por:
 - a) Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
 - b) Las plataformas ofrecen experiencias a la medida, permitiendo a los usuarios personalizar sus perfiles, recibir recomendaciones ajustadas al contenido y participar en filtrado colaborativo.
 - c) La existencia de sitios web dinámicos e interactivos que permiten a los usuarios participar, comentar e interactuar tanto con los creadores de contenido como con otros usuarios.
 - d) El uso de distintas tecnologías para crear experiencias web más interactivas y con mayor capacidad de respuesta.
2. Al realizar la optimización de contenido para SEO debe considerarse:
 - a) Utilizar etiquetas de título y meta descripciones únicas y relevantes para cada página.
 - b) Seleccionar las palabras clave al azar.
 - c) Crear contenido valioso y original que satisfaga las necesidades de los usuarios.
 - d) Incluir palabras clave de manera excesiva en el contenido para mejorar el posicionamiento.
3. Si se tiene el conjunto de páginas interconectadas $\{A \rightarrow B, C; B \rightarrow C; C \rightarrow A\}$, entonces la página con valor más alto de PageRank es:
 - a) B.
 - b) C.
 - c) A y B.
 - d) A.
4. El algoritmo Hypertext Induced Topic Selection (HITS) intenta buscar nodos especiales. Estos son conocidos como:
 - a) Sensibles.
 - b) Centrales.
 - c) Hubs.
 - d) Autoridades.
5. La centralidad de intermediación de un nodo indica:
 - a) La frecuencia con la que un nodo actúa como puente en el camino más corto entre otros dos nodos.
 - b) El número total de conexiones entrantes y salientes.
 - c) La resistencia del nodo a fallos.
 - d) La cantidad de nodos vecinos directos.
6. Un SRI es capaz de:
 - a) No necesitar de ningún almacenamiento externo para alojar los índices de los datos.
 - b) Reducir el tiempo de indexación de los datos si utiliza servidores distribuidos dentro de la red para que cada uno ejecute la indexación del mismo conjunto de datos.
 - c) Generar índices invertidos de manera óptima sin considerar el contexto.
 - d) Crear los índices asociados a los datos sin tener que analizar cada dato.
7. El concepto de *rank sink* en el algoritmo de PageRank representa páginas web:
 - a) Con una baja calidad de contenido y una cantidad insuficiente de enlaces salientes, lo que las hace menos relevantes en los resultados de búsqueda.
 - b) Con distribuciones uniformes de la puntuación de PageRank.
 - c) Con un alto número de enlaces salientes que no reciben enlaces entrantes, lo que puede afectar negativamente su puntuación de PageRank.
 - d) Que tienen una puntuación de PageRank más alta que otras debido a la manipulación de enlaces entrantes y salientes.
8. Una plataforma de comercio electrónico desea mejorar la experiencia del usuario al permitir una navegación más personalizada y contextualizada. Actualmente los usuarios tienen dificultades para encontrar productos relevantes debido a la gran cantidad de opciones disponibles. La empresa está interesada en implementar características de la Web

- 2.5 y la Web Semántica para abordar este problema. ¿Qué características podrían ayudar para ofrecer una navegación más personalizada y contextualizada?
- La implementación de ontologías y metadatos para enriquecer la descripción de productos y mejorar la precisión de las recomendaciones.
 - La optimización de la velocidad de carga del sitio web para mejorar la experiencia del usuario y reducir el abandono del carrito de compra.
 - La incorporación de los productos en tendencia en el mercado.
 - La integración de redes sociales para permitir la recomendación de productos basada en las preferencias de amigos y contactos.
9. En la RI en el contexto de Big Data, se puede asegurar que:
- La indexación distribuida es una técnica obsoleta.
 - La indexación distribuida almacena todos los datos en un solo servidor para facilitar su acceso y búsqueda.
 - La indexación distribuida divide los datos en múltiples fragmentos que se almacenan en varios nodos para permitir búsquedas paralelas y mejorar la escalabilidad.
 - La indexación distribuida no ofrece ventajas en términos de rendimiento y escalabilidad en comparación con la indexación centralizada.
10. Al “relajar” el concepto de clique en la detección de comunidades se intenta solucionar:
- La uniformidad de los nodos en términos de grado.
 - La necesidad de datos externos para analizar la red.
 - La imposibilidad de conectar nodos distantes.
 - El solapamiento y la complejidad computacional.
11. La diferencia entre la indexación por tokens y la indexación por conceptos puede definirse como:
- La indexación por tokens normaliza los datos reduciéndolos a su forma básica, mientras que la indexación por conceptos utiliza un método de ordenación para organizar los términos característicos de los datos.
 - La indexación por tokens asigna un valor numérico a cada término de los datos, mientras que la indexación por conceptos utiliza algoritmos de encriptación para proteger la privacidad de los datos.
 - La indexación por tokens asigna pesos a los términos basados en su importancia relativa, mientras que la indexación por conceptos utiliza un sistema de etiquetado para asociar términos con características generales.
 - La indexación por tokens divide los datos en términos individuales, mientras que la indexación por conceptos agrupa los datos en categorías definidas.
12. No se considera como técnica para detectar comunidades en una red:
- Analizar la mutualidad de los enlaces.
 - Usar el agrupamiento jerárquico.
 - Utilizar el algoritmo de K-Means.
 - Encontrar cliques de vértices de grado par.
13. El propósito de la política de revisitado en los Web Crawlers es:
- Determinar la frecuencia con la que los crawlers deben volver a visitar una URL específica para mantener la información actualizada en el índice de búsqueda.
 - Limitar el acceso de los crawlers a ciertos servidores luego de visitar las páginas alojadas en estos.
 - Establecer reglas sobre el tiempo máximo que los crawlers pueden pasar en un sitio web durante cada visita para evitar sobrecargar los servidores.
 - Definir la prioridad de rastreo de las URLs, determinando el orden en que los crawlers visitan y procesan cada página web.
14. En la RI en el contexto de Big Data, se puede asegurar que:
- El almacenamiento distribuido es una técnica obsoleta.
 - El almacenamiento distribuido centraliza todos los datos en un único servidor para facilitar su gestión y mantenimiento.
 - El almacenamiento distribuido reparte los datos en múltiples servidores para mejorar la disponibilidad y la redundancia del sistema.
 - El almacenamiento distribuido ofrece ventajas significativas en términos de escalabilidad y rendimiento en comparación con el almacenamiento centralizado.
15. La afirmación que mejor describe la política de amabilidad en los Web Crawlers es:
- Los Web Crawlers son libres de recopilar datos de cualquier sitio web sin restricciones, independientemente de la cantidad de tráfico que generen.

- b) La política de amabilidad establece pautas y reglas sobre cómo los crawlers deben interactuar con los sitios web para minimizar la carga del servidor y respetar las directivas de los administradores del sitio.
- c) La política de amabilidad de los Web Crawlers dicta que los crawlers deben priorizar ciertos tipos de contenido sobre otros, ignorando completamente ciertas páginas web.
- d) Los crawlers se diseñan para acceder a sitios web sin restricciones y extraer datos de manera agresiva para su indexación.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. La afirmación que mejor describe la política de amabilidad en los Web Crawlers es:
 - a) Los Web Crawlers son libres de recopilar datos de cualquier sitio web sin restricciones, independientemente de la cantidad de tráfico que generen.
 - b) La política de amabilidad establece pautas y reglas sobre cómo los crawlers deben interactuar con los sitios web para minimizar la carga del servidor y respetar las directivas de los administradores del sitio.
 - c) La política de amabilidad de los Web Crawlers dicta que los crawlers deben priorizar ciertos tipos de contenido sobre otros, ignorando completamente ciertas páginas web.
 - d) Los crawlers se diseñan para acceder a sitios web sin restricciones y extraer datos de manera agresiva para su indexación.
2. Se puede afirmar que:
 - a) La cantidad de conexiones de un nodo siempre indica su influencia en la red.
 - b) Todas las relaciones en una red tienen la misma importancia para el análisis.
 - c) No es posible identificar subgrupos dentro de una red utilizando análisis de redes.
 - d) El tamaño de una red es siempre indicativo de su efectividad en la transmisión de información.
3. ¿Cuál de las siguientes opciones describe mejor la diferencia clave entre Web Crawling y Web Scraping?
 - a) Web Scraping implica el análisis de la estructura y el contenido de las páginas web para extraer datos, mientras que Web Crawling se refiere a la descarga y almacenamiento de páginas web completas.
 - b) Web Scraping es más eficaz para rastrear e indexar contenido web para motores de búsqueda, mientras que Web Crawling se utiliza principalmente para la extracción de datos en proyectos de investigación.
 - c) Web Crawling se centra en la exploración y recopilación de enlaces de múltiples sitios web, mientras que Web Scraping se enfoca en la extracción específica de datos de páginas web individuales.
 - d) Web Crawling se realiza utilizando herramientas de automatización como Selenium WebDriver, mientras que Web Scraping se lleva a cabo mediante el análisis de HTML y CSS.
4. La web actual se enfrenta a problemas como:
 - a) Presencia de una alta calidad en los datos.
 - b) Heterogeneidad en los datos.
 - c) Presencia de datos volátiles y distribuidos.
 - d) Presencia de grandes volúmenes de datos estructurados.
5. En una empresa comercial, se tiene una red donde los nodos corresponden a los empleados y las aristas representan la frecuencia con la que colaboran en las ventas. Se busca mejorar la colaboración entre los empleados para aumentar las ventas totales, por lo que la directiva debe:
 - a) No tener en cuenta la centralidad de cercanía de los empleados al diseñar estrategias de colaboración, ya que esto podría introducir complicaciones adicionales en el proceso de trabajo.
 - b) Utilizar la centralidad de cercanía para identificar los empleados menos cercanos a otros en la red y asignarles tareas individuales para evitar posibles conflictos y desacuerdos en el proceso de colaboración.
 - c) No considerar la centralidad de cercanía de los empleados al diseñar estrategias de colaboración, ya que todos los empleados tienen el mismo acceso a la información y los recursos.
 - d) Utilizar la centralidad de cercanía para identificar los empleados más cercanos a otros en la red y promover la colaboración entre ellos, facilitando así la comunicación y el intercambio de conocimientos para mejorar las ventas.
6. Si se tiene el conjunto de páginas interconectadas $\{A \rightarrow B, C; B \rightarrow C; C \rightarrow A\}$, entonces la página con valor más alto de PageRank es:
 - a) B.
 - b) C.

- c) A y B.
d) A.
7. Si una red cumple la propiedad de ser un grafo de mundo pequeño, entonces se conoce que:
- La red posee pocos vértices.
 - La longitud media del camino entre todo par de vértices es pequeña.
 - El número de componentes fuertemente conexas está relacionado con la cantidad de grafos K_n presentes.
 - El grafo es un anillo regular de grado 5.
8. En una red de transporte donde cada nodo es una parada de autobús y las aristas representan si existe un carro que pasa por ambos sitios, ¿qué puede mejorar el sistema de transporte?
- Utilizar la centralidad de intermediación para identificar las paradas de transferencia clave y establecer nuevas conexiones entre diferentes líneas de transporte público.
 - Utilizar el grafo inducido de los nodos con mayor valor en la centralidad de grado para aplicar la centralidad de intermediación con el propósito de reforzar las paradas con mayor tráfico.
 - No implementar sistemas de información en tiempo real para los usuarios, ya que pueden aumentar la carga de los trabajadores.
 - No considerar la centralidad de intermediación de las estaciones de transporte público al planificar rutas y horarios, ya que no tiene impacto en las conexiones entre las paradas.
9. En un SRI la indexación:
- Mejora la experiencia del usuario.
 - Consiste en asociar un identificador único a cada dato almacenado en el sistema.
 - Facilita la RI relevante.
 - Permite la organización y la categorización de la información.
10. En la RI en el contexto de Big Data, se puede asegurar que:
- La indexación distribuida es una técnica obsoleta.
 - La indexación distribuida almacena todos los datos en un solo servidor para facilitar su acceso y búsqueda.
 - La indexación distribuida divide los datos en múltiples fragmentos que se almacenan en varios nodos para permitir búsquedas paralelas y mejorar la escalabilidad.
 - La indexación distribuida no ofrece ventajas en términos de rendimiento y escalabilidad en comparación con la indexación centralizada.
11. La diferencia entre la indexación por tokens y la indexación por conceptos puede definirse como:
- La indexación por tokens normaliza los datos reduciéndolos a su forma básica, mientras que la indexación por conceptos utiliza un método de ordenación para organizar los términos característicos de los datos.
 - La indexación por tokens asigna un valor numérico a cada término de los datos, mientras que la indexación por conceptos utiliza algoritmos de encriptación para proteger la privacidad de los datos.
 - La indexación por tokens asigna pesos a los términos basados en su importancia relativa, mientras que la indexación por conceptos utiliza un sistema de etiquetado para asociar términos con características generales.
 - La indexación por tokens divide los datos en términos individuales, mientras que la indexación por conceptos agrupa los datos en categorías definidas.
12. La Web 1.0 se caracteriza por:
- Los propietarios de los sitios web proporcionan contenido de forma periódica.
 - Uso de comunidades virtuales para popularizar los sitios web de noticias.
 - Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
 - Los sitios web se centran en brindar información en lugar de facilitar la colaboración o participación de los usuarios.
13. El algoritmo de PageRank puede describirse como un procedimiento utilizado para:
- Determinar la velocidad de carga de una página web en un navegador.
 - Calcular la relevancia de una página web en función de la cantidad y calidad de los enlaces que apuntan hacia ella.
 - Clasificar las páginas web en función de su edad y autoridad.
 - Prevenir el spam y el contenido no deseado en las páginas web.
14. Una empresa de comercio electrónico necesita mejorar su motor de búsqueda para proporcionar resultados más relevantes a sus usuarios. Actualmente, los resultados de la búsqueda no son precisos y los usuarios a menudo encuentran dificultades para

encontrar productos específicos. La empresa está considerando implementar una indexación por conceptos para mejorar la relevancia de los resultados de búsqueda. ¿Qué beneficios podría provocar este cambio?

- a) Permite adaptarse fácilmente a cambios en el vocabulario y la terminología utilizada en las descripciones de los productos.
 - b) Facilita la visualización de productos al agruparlos por categorías o características comunes.
 - c) Mejora la precisión en la búsqueda de productos relacionados, incluso cuando no coinciden exactamente con los términos de búsqueda del usuario.
 - d) Ayuda a identificar automáticamente relaciones entre productos, lo que puede mejorar las recomendaciones personalizadas a los usuarios.
15. Considerando las dimensiones y desafíos inherentes a Big Data puede afirmarse, tomando en cuenta las características clave y las implicaciones para su procesamiento y análisis, que:
- a) Big Data se caracteriza principalmente por su pequeño volumen y uniformidad, permitiendo un procesamiento eficiente con mínimas adaptaciones de las herramientas de análisis de datos tradicionales.
 - b) Big Data no desafía la capacidad de las herramientas tradicionales de procesamiento de datos para capturar, almacenar, gestionar y analizar efectivamente la información, dada la evolución constante de las capacidades computacionales y algoritmos de optimización.
 - c) Aunque Big Data puede incluir datos estructurados, su naturaleza se expande al incorporar grandes cantidades de datos no estructurados y semiestructurados, lo que exige el uso de tecnologías especializadas en almacenamiento y procesamiento como Hadoop y sistemas de bases de datos NoSQL.
 - d) Además de su complejidad y diversidad, Big Data introduce desafíos significativos en términos de veracidad y variabilidad al requerir métodos avanzados de limpieza y validación de datos para asegurar la integridad del análisis.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. El posicionamiento de un sitio web en los motores de búsqueda puede ser afectado por:
 - a) Mantener una estructura de URL clara y coherente.
 - b) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - c) Contenido sin valor en el sitio web.
 - d) Obtener enlaces de sitios web irrelevantes y de baja calidad.
2. La afirmación que mejor refleja el principio subyacente de PageRank, considerando su importancia en la clasificación de los sitios web, es:
 - a) El algoritmo de PageRank considera tanto la cantidad como la calidad de los enlaces entrantes, asignando mayor valor a los enlaces provenientes de sitios web considerados como “importantes”.
 - b) PageRank valora más la cantidad de enlaces entrantes a una página web, independientemente de la calidad o relevancia de estos enlaces.
 - c) La efectividad de PageRank se basa exclusivamente en el análisis de las palabras clave contenidas en los enlaces entrantes, sin considerar la estructura de enlace de la Web.
 - d) PageRank opera bajo el supuesto de que los enlaces entrantes y salientes tienen el mismo impacto en la valoración de la relevancia de una página web.
3. Una buena práctica de SEO para mejorar el posicionamiento de un sitio web en los motores de búsqueda es:
 - a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Copiar contenido directamente de otros sitios web populares para aumentar la cantidad de páginas indexadas.
 - c) Llenar el contenido con palabras clave irrelevantes para aumentar la densidad de palabras clave.
 - d) Obtener enlaces de otros sitios web relevantes y de calidad que apunten al sitio.
4. El algoritmo de PageRank puede describirse como un procedimiento utilizado para:
 - a) Determinar la velocidad de carga de una página web en un navegador.
 - b) Calcular la relevancia de una página web en función de la cantidad y calidad de los enlaces que apuntan hacia ella.
 - c) Clasificar las páginas web en función de su edad y autoridad.
 - d) Prevenir el spam y el contenido no deseado en las páginas web.
5. Un sistema cuenta con la siguiente información:
 - Juan pertenece a la marina.
 - Juan es capellán (sacerdote encargado de una tarea específica fuera de la parroquia).
 - Los infantes de la marina suelen ser bebedores de cerveza.
 - Un capellán no suele ser bebedor de cerveza.
 - Un bebedor de cerveza suele tener sobrepeso.
 - Por lo general, un infante de la marina está en buena condición física.

Si se representa la información en una red de herencia se puede concluir que:

 - a) Existen dos razonamientos cancelables.
 - b) Juan tiene sobrepeso.
 - c) No se puede asegurar que Juan sea capellán.
 - d) No se puede asegurar que Juan esté en buena condición física.
6. Considerando las dimensiones y desafíos inherentes a Big Data puede afirmarse, tomando en cuenta las características clave y las implicaciones para su procesamiento y análisis, que:
 - a) Big Data se caracteriza principalmente por su pequeño volumen y uniformidad, permitiendo un procesamiento eficiente con mínimas adaptaciones de las herramientas de análisis de datos tradicionales.
 - b) Big Data no desafía la capacidad de las herramientas tradicionales de procesamiento de datos para capturar, almacenar, gestionar y analizar efectivamente la información, dada la evolución constante de las capacidades computacionales y algoritmos de optimización.

- c) Aunque Big Data puede incluir datos estructurados, su naturaleza se expande al incorporar grandes cantidades de datos no estructurados y semiestructurados, lo que exige el uso de tecnologías especializadas en almacenamiento y procesamiento como Hadoop y sistemas de bases de datos NoSQL.
- d) Además de su complejidad y diversidad, Big Data introduce desafíos significativos en términos de veracidad y variabilidad al requerir métodos avanzados de limpieza y validación de datos para asegurar la integridad del análisis.
7. La transición de la Web 1.0 a la Web 2.0 se caracterizó principalmente por:
- La reducción en el uso de HTML y CSS en el desarrollo de sitios web.
 - La disminución de la importancia de los motores de búsqueda en la navegación web.
 - El cambio de páginas web estáticas a dinámicas, permitiendo la interacción del usuario y la generación de contenido.
 - El aumento en la velocidad de conexión a internet, que permitió una mejor calidad de las páginas web.
8. ¿Qué algoritmos permiten obtener información de una red?
- Índices de centralidad.
 - Detección de comunidades.
 - Hypertext Induced Topic Selection (HITS).
 - PageRank.
9. En el algoritmo de Indexación Basada en Clasificación Bloqueada (BSBI), ¿cuál es el paso final para crear un índice invertido para la colección completa de los datos?
- Dividir la colección de datos en bloques de tamaño fijo.
 - Eliminar los términos duplicados de los índices.
 - Indexar cada bloque de forma independiente.
 - Fusionar los índices invertidos de cada bloque.
10. No se considera como técnica para detectar comunidades en una red:
- Analizar la mutualidad de los enlaces.
 - Usar el agrupamiento jerárquico.
 - Utilizar el algoritmo de K-Means.
 - Encontrar cliques de vértices de grado par.
11. La web actual se enfrenta a problemas como:
- Presencia de una alta calidad en los datos.
 - Heterogeneidad en los datos.
 - Presencia de datos volátiles y distribuidos.
 - Presencia de grandes volúmenes de datos estructurados.
12. Si se tiene el conjunto de páginas interconectadas $\{A \rightarrow B, C; B \rightarrow C; C \rightarrow A\}$, entonces la página con valor más alto de PageRank es:
- B.
 - C.
 - A y B.
 - A.
13. Una empresa de comercio electrónico necesita mejorar su motor de búsqueda para proporcionar resultados más relevantes a sus usuarios. Actualmente, los resultados de la búsqueda no son precisos y los usuarios a menudo encuentran dificultades para encontrar productos específicos. La empresa está considerando implementar una indexación por conceptos para mejorar la relevancia de los resultados de búsqueda. ¿Qué beneficios podría provocar este cambio?
- Permite adaptarse fácilmente a cambios en el vocabulario y la terminología utilizada en las descripciones de los productos.
 - Facilita la visualización de productos al agruparlos por categorías o características comunes.
 - Mejora la precisión en la búsqueda de productos relacionados, incluso cuando no coinciden exactamente con los términos de búsqueda del usuario.
 - Ayuda a identificar automáticamente relaciones entre productos, lo que puede mejorar las recomendaciones personalizadas a los usuarios.
14. El componente responsable de la gestión de recursos y planificación de tareas en Hadoop es:
- MapReduce.
 - HDFS.
 - Hadoop Common.
 - YARN.
15. En una empresa comercial, se tiene una red donde los nodos corresponden a los empleados y las aristas representan la frecuencia con la que colaboran en las ventas. Se busca mejorar la colaboración entre los empleados para aumentar las ventas totales, por lo que la directiva debe:
- No tener en cuenta la centralidad de cercanía de los empleados al diseñar estrategias de colaboración, ya que esto podría introducir complicaciones adicionales en el proceso de trabajo.

- b) Utilizar la centralidad de cercanía para identificar los empleados menos cercanos a otros en la red y asignarles tareas individuales para evitar posibles conflictos y desacuerdos en el proceso de colaboración.
- c) No considerar la centralidad de cercanía de los empleados al diseñar estrategias de colaboración, ya que todos los empleados tienen el mismo acceso a la información y los recursos.
- d) Utilizar la centralidad de cercanía para identificar los empleados más cercanos a otros en la red y promover la colaboración entre ellos, facilitando así la comunicación y el intercambio de conocimientos para mejorar las ventas.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Con respecto a WordNet y su contribución a las ontologías y la representación del conocimiento en los sistemas de RI, puede afirmarse que:
 - a) En WordNet, una palabra está asociada a un conjunto de sinónimos (synsets), siendo estas palabras intercambiables en un contexto.
 - b) WordNet diferencia claramente entre los significados de palabras según su uso en diferentes contextos, lo que permite aplicaciones avanzadas en desambiguación semántica más allá de los sistemas de recomendación y RI.
 - c) La integración de WordNet en sistemas de RI permite la expansión de consultas y la mejora de la precisión de los resultados al entender mejor el significado de los términos de búsqueda a través de su contexto semántico y las relaciones entre palabras.
 - d) Una limitación de WordNet en la representación del conocimiento es su enfoque en el idioma inglés, lo que plantea desafíos en la aplicación global y la interoperabilidad con sistemas de información multilingües.
2. ¿Qué es un 'Uniform Resource Locator (URL) Frontier' en el contexto de Web Crawling?
 - a) Una base de datos que almacena URLs únicas identificadas como recursos en la web.
 - b) Una técnica para filtrar URLs irrelevantes y mejorar la eficiencia del crawling.
 - c) Un protocolo que define cómo se deben formatear las URLs para el crawling.
 - d) Una lista prioritaria de URLs que aún no han sido visitadas por el crawler.
3. Una empresa de comercio electrónico necesita mejorar su motor de búsqueda para proporcionar resultados más relevantes a sus usuarios. Actualmente, los resultados de la búsqueda no son precisos y los usuarios a menudo encuentran dificultades para encontrar productos específicos. La empresa está considerando implementar una indexación por conceptos para mejorar la relevancia de los resultados de búsqueda. ¿Qué beneficios podría provocar este cambio?
 - a) Permite adaptarse fácilmente a cambios en el vocabulario y la terminología utilizada en las descripciones de los productos.
 - b) Facilita la visualización de productos al agruparlos por categorías o características comunes.
 - c) Mejora la precisión en la búsqueda de productos relacionados, incluso cuando no coinciden exactamente con los términos de búsqueda del usuario.
 - d) Ayuda a identificar automáticamente relaciones entre productos, lo que puede mejorar las recomendaciones personalizadas a los usuarios.
4. La premisa básica del algoritmo de PageRank para clasificar páginas web en los resultados de búsqueda es:
 - a) Los enlaces entrantes a una página web desde otras páginas contribuyen a su importancia y clasificación.
 - b) La antigüedad de la página web es el principal factor para determinar su clasificación.
 - c) La cantidad de visitas que recibe una página web determina su clasificación.
 - d) El contenido y la relevancia de las palabras clave en la página web determinan su posición.
5. El posicionamiento de un sitio web en los motores de búsqueda puede ser afectado por:
 - a) Mantener una estructura de URL clara y coherente.
 - b) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - c) Contenido sin valor en el sitio web.
 - d) Obtener enlaces de sitios web irrelevantes y de baja calidad.
6. El algoritmo de PageRank puede describirse como un procedimiento utilizado para:
 - a) Determinar la velocidad de carga de una página web en un navegador.
 - b) Calcular la relevancia de una página web en función de la cantidad y calidad de los enlaces que apuntan hacia ella.

- c) Clasificar las páginas web en función de su edad y autoridad.
 - d) Prevenir el spam y el contenido no deseado en las páginas web.
7. La computación evolutiva:
- a) Utiliza algoritmos para buscar soluciones óptimas en grandes espacios de búsqueda, lo que la hace adecuada para problemas de optimización en la RI.
 - b) Es útil solo para problemas de clasificación de documentos y no para otras tareas de RI en general.
 - c) No es aplicable en la RI debido a la complejidad de los algoritmos evolutivos.
 - d) Solo puede manejar conjuntos de datos pequeños y no es escalable a grandes volúmenes de datos.
8. Se puede afirmar que:
- a) La cantidad de conexiones de un nodo siempre indica su influencia en la red.
 - b) Todas las relaciones en una red tienen la misma importancia para el análisis.
 - c) No es posible identificar subgrupos dentro de una red utilizando análisis de redes.
 - d) El tamaño de una red es siempre indicativo de su efectividad en la transmisión de información.
9. En el diseño de un videojuego de roles (RPG) en el que los personajes tienen atributos como salud, fuerza y velocidad, ¿cuál de las siguientes opciones representa mejor una implementación de la representación del conocimiento orientado a objetos?
- a) Cada personaje se representa como un objeto con propiedades como salud, fuerza y velocidad, y métodos para modificar y consultar estos valores.
 - b) Cada personaje se representa como una función que calcula sus atributos en función de su nivel y experiencia.
 - c) Cada personaje se representa como una matriz de números que almacena sus valores de atributos.
 - d) Cada personaje se representa como una lista de cadenas de texto que describen sus características físicas y habilidades.
10. En un sistema de control de tráfico urbano basado en reglas, ¿cuál de las siguientes reglas sería más efectiva para manejar situaciones de congestión en una intersección?
- a) Si hay pocos vehículos en la intersección, reducir el tiempo de los semáforos en verde.
 - b) Si hay muchos vehículos en la intersección, aumentar el tiempo de los semáforos en verde.
 - c) Si hay un vehículo de emergencia en la intersección, detener todos los demás vehículos.
 - d) Si hay muchos peatones cruzando la intersección, reducir el tiempo de los semáforos en rojo.
11. Un SRI es capaz de:
- a) No necesitar de ningún almacenamiento externo para alojar los índices de los datos.
 - b) Reducir el tiempo de indexación de los datos si utiliza servidores distribuidos dentro de la red para que cada uno ejecute la indexación del mismo conjunto de datos.
 - c) Generar índices invertidos de manera óptima sin considerar el contexto.
 - d) Crear los índices asociados a los datos sin tener que analizar cada dato.
12. Considerando las dimensiones y desafíos inherentes a Big Data puede afirmarse, tomando en cuenta las características clave y las implicaciones para su procesamiento y análisis, que:
- a) Big Data se caracteriza principalmente por su pequeño volumen y uniformidad, permitiendo un procesamiento eficiente con mínimas adaptaciones de las herramientas de análisis de datos tradicionales.
 - b) Big Data no desafía la capacidad de las herramientas tradicionales de procesamiento de datos para capturar, almacenar, gestionar y analizar efectivamente la información, dada la evolución constante de las capacidades computacionales y algoritmos de optimización.
 - c) Aunque Big Data puede incluir datos estructurados, su naturaleza se expande al incorporar grandes cantidades de datos no estructurados y semiestructurados, lo que exige el uso de tecnologías especializadas en almacenamiento y procesamiento como Hadoop y sistemas de bases de datos NoSQL.
 - d) Además de su complejidad y diversidad, Big Data introduce desafíos significativos en términos de veracidad y variabilidad al requerir métodos avanzados de limpieza y validación de datos para asegurar la integridad del análisis.
13. La integración de Hadoop y MapReduce en la RI trae como ventaja:
- a) La posibilidad de la extracción de información relevante y la generación de resultados significativos de grandes conjuntos de datos.
 - b) La reducción de los costos operativos a cero.
 - c) La eliminación de la necesidad de sistemas de bases de datos.

- d) La garantía de la privacidad absoluta de los datos procesados.
14. Una plataforma de comercio electrónico desea mejorar la experiencia del usuario al permitir una navegación más personalizada y contextualizada. Actualmente los usuarios tienen dificultades para encontrar productos relevantes debido a la gran cantidad de opciones disponibles. La empresa está interesada en implementar características de la Web 2.5 y la Web Semántica para abordar este problema. ¿Qué características podrían ayudar para ofrecer una navegación más personalizada y contextualizada?
- a) La implementación de ontologías y metadatos para enriquecer la descripción de productos y mejorar la precisión de las recomendaciones.
 - b) La optimización de la velocidad de carga del sitio web para mejorar la experiencia del usuario y reducir el abandono del carrito de compra.
 - c) La incorporación de los productos en tendencia en el mercado.
- d) La integración de redes sociales para permitir la recomendación de productos basada en las preferencias de amigos y contactos.
15. Se tiene un grafo donde cada nodo es un personaje de cierto libro de cuentos y la existencia de las aristas está definida si dos personajes aparecen en el mismo cuento. Se puede asegurar que:
- a) La centralidad de grado es útil para identificar los personajes con más conexiones dentro de la red, lo que puede indicar su importancia en el libro.
 - b) La centralidad de cercanía indica el grado de conexión de cada personaje con el resto de los personajes de los cuentos del libro.
 - c) La centralidad de intermediación identifica a los personajes que actúan como conectores entre personajes de cuentos distintos.
 - d) La centralidad de cercanía ofrece una relación entre la cantidad de vecinos de un nodo con respecto a la longitud máxima de un camino dentro del grafo partiendo del nodo en cuestión.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Al “relajar” el concepto de clique en la detección de comunidades se intenta solucionar:
 - a) La uniformidad de los nodos en términos de grado.
 - b) La necesidad de datos externos para analizar la red.
 - c) La imposibilidad de conectar nodos distantes.
 - d) El solapamiento y la complejidad computacional.
2. Considerando las prácticas éticas y legales en el Web Scraping, se puede asegurar que:
 - a) Web Scraping sobre datos personales sin consentimiento es generalmente aceptado si los datos se utilizan con fines de investigación.
 - b) Es importante revisar y respetar los términos de servicio del sitio web, así como las leyes aplicables de protección de datos y derechos de autor, antes de realizar Web Scraping.
 - c) Web Scraping requiere considerar las políticas de **robots.txt** del sitio web objetivo.
 - d) La extracción de datos mediante Web Scraping siempre es legal, independientemente de las leyes locales sobre derechos de autor y privacidad de datos definidos en los sitios web.
3. No se considera como técnica para detectar comunidades en una red:
 - a) Analizar la mutualidad de los enlaces.
 - b) Usar el agrupamiento jerárquico.
 - c) Utilizar el algoritmo de K-Means.
 - d) Encontrar cliques de vértices de grado par.
4. En un sistema de control de tráfico urbano basado en reglas, ¿cuál de las siguientes reglas sería más efectiva para manejar situaciones de congestión en una intersección?
 - a) Si hay pocos vehículos en la intersección, reducir el tiempo de los semáforos en verde.
 - b) Si hay muchos vehículos en la intersección, aumentar el tiempo de los semáforos en verde.
 - c) Si hay un vehículo de emergencia en la intersección, detener todos los demás vehículos.
 - d) Si hay muchos peatones cruzando la intersección, reducir el tiempo de los semáforos en rojo.
5. La afirmación que mejor refleja el principio subyacente de PageRank, considerando su importancia en la clasificación de los sitios web, es:
 - a) El algoritmo de PageRank considera tanto la cantidad como la calidad de los enlaces entrantes, asignando mayor valor a los enlaces provenientes de sitios web considerados como “importantes”.
 - b) PageRank valora más la cantidad de enlaces entrantes a una página web, independientemente de la calidad o relevancia de estos enlaces.
 - c) La efectividad de PageRank se basa exclusivamente en el análisis de las palabras clave contenidas en los enlaces entrantes, sin considerar la estructura de enlace de la Web.
 - d) PageRank opera bajo el supuesto de que los enlaces entrantes y salientes tienen el mismo impacto en la valoración de la relevancia de una página web.
6. Los Web Crawlers se enfrentan a desafíos constantes. Dentro de ellos se encuentran:
 - a) La falta de acceso a la base de datos del servidor web para extraer información actualizada.
 - b) La modificación del código y la estructura del sitio web.
 - c) La incapacidad para interpretar correctamente el lenguaje de programación utilizado en el desarrollo de los sitios web.
 - d) La dificultad para generar el contenido dinámico en tiempo real.
7. Se tiene una red de ingredientes donde cada uno representa un nodo y las aristas simbolizan que los ingredientes forman parte de una misma receta. Se busca mejorar la experiencia culinaria mediante la elaboración de combinaciones de ingredientes más interesantes y creativas, para lo cual se debe:
 - a) No tener en cuenta la centralidad de grado de los ingredientes para no darle mayor importancia a los ingredientes más comunes.

- b) No considerar la centralidad de grado de los ingredientes, ya que todas las combinaciones de ingredientes son igualmente válidas.
 - c) Utilizar la centralidad de grado para identificar los ingredientes más populares en la red y crear combinaciones que incluyan una variedad de ingredientes menos comunes.
 - d) Utilizar la centralidad de grado para identificar los ingredientes menos conectados en la red y tomarlos en cuenta para su inclusión en las combinaciones.
8. El posicionamiento de un sitio web en los motores de búsqueda puede ser afectado por:
- a) Mantener una estructura de URL clara y coherente.
 - b) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - c) Contenido sin valor en el sitio web.
 - d) Obtener enlaces de sitios web irrelevantes y de baja calidad.
9. Dentro del ecosistema de Hadoop, el HDFS se caracteriza por:
- a) La capacidad ilimitada de almacenamiento.
 - b) La tolerancia a fallos mediante la replicación de datos.
 - c) El almacenamiento exclusivo para archivos de texto.
 - d) El modelo de acceso y de escritura de datos en tiempo real.
10. El conjunto frontera de URLs en un Web Crawler:
- a) Delimita los sitios web que visitará en futuras iteraciones del proceso.
 - b) Indica las secciones que no pueden ser visitadas de cada sitio web.
 - c) Almacena hipervínculos que pertenecen al mismo dominio del conjunto semilla de URLs con que inició el crawler.
 - d) Es similar a un conjunto de URLs que esperan ser visitadas.
11. En un SRI la indexación:
- a) Mejora la experiencia del usuario.
 - b) Consiste en asociar un identificador único a cada dato almacenado en el sistema.
 - c) Facilita la RI relevante.
 - d) Permite la organización y la categorización de la información.
12. Se puede afirmar que:
- a) Uno de los desafíos en el procesamiento de Big Data es la capacidad de gestionar y analizar datos provenientes de diversas fuentes y en diferentes formatos de manera eficiente.
 - b) MapReduce es un modelo de procesamiento distribuido utilizado para trabajar con grandes volúmenes de datos.
 - c) Los SRI pueden beneficiarse de MapReduce para mejorar la RI relevante.
 - d) El término “Big Data” se refiere exclusivamente al volumen de datos que una organización maneja, sin tener en cuenta la velocidad, la variedad y la veracidad de los datos.
13. En una red de transporte donde cada nodo es una parada de autobús y las aristas representan si existe un carro que pasa por ambos sitios, ¿qué puede mejorar el sistema de transporte?
- a) Utilizar la centralidad de intermediación para identificar las paradas de transferencia clave y establecer nuevas conexiones entre diferentes líneas de transporte público.
 - b) Utilizar el grafo inducido de los nodos con mayor valor en la centralidad de grado para aplicar la centralidad de intermediación con el propósito de reforzar las paradas con mayor tráfico.
 - c) No implementar sistemas de información en tiempo real para los usuarios, ya que pueden aumentar la carga de los trabajadores.
 - d) No considerar la centralidad de intermediación de las estaciones de transporte público al planificar rutas y horarios, ya que no tiene impacto en las conexiones entre las paradas.
14. Una plataforma de comercio electrónico desea mejorar la experiencia del usuario al permitir una navegación más personalizada y contextualizada. Actualmente los usuarios tienen dificultades para encontrar productos relevantes debido a la gran cantidad de opciones disponibles. La empresa está interesada en implementar características de la Web 2.5 y la Web Semántica para abordar este problema. ¿Qué características podrían ayudar para ofrecer una navegación más personalizada y contextualizada?
- a) La implementación de ontologías y metadatos para enriquecer la descripción de productos y mejorar la precisión de las recomendaciones.
 - b) La optimización de la velocidad de carga del sitio web para mejorar la experiencia del usuario y reducir el abandono del carrito de compra.
 - c) La incorporación de los productos en tendencia en el mercado.
 - d) La integración de redes sociales para permitir la recomendación de productos basada en las preferencias de amigos y contactos.

15. La diferencia entre la indexación por tokens y la indexación por conceptos puede definirse como:

- a) La indexación por tokens normaliza los datos reduciéndolos a su forma básica, mientras que la indexación por conceptos utiliza un método de ordenación para organizar los términos característicos de los datos.
- b) La indexación por tokens asigna un valor numérico a cada término de los datos, mientras que la indexación por conceptos utiliza algoritmos de encriptación para proteger la privacidad

de los datos.

- c) La indexación por tokens asigna pesos a los términos basados en su importancia relativa, mientras que la indexación por conceptos utiliza un sistema de etiquetado para asociar términos con características generales.
- d) La indexación por tokens divide los datos en términos individuales, mientras que la indexación por conceptos agrupa los datos en categorías definidas.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. El concepto de *rank sink* en el algoritmo de PageRank representa páginas web:
 - a) Con una baja calidad de contenido y una cantidad insuficiente de enlaces salientes, lo que las hace menos relevantes en los resultados de búsqueda.
 - b) Con distribuciones uniformes de la puntuación de PageRank.
 - c) Con un alto número de enlaces salientes que no reciben enlaces entrantes, lo que puede afectar negativamente su puntuación de PageRank.
 - d) Que tienen una puntuación de PageRank más alta que otras debido a la manipulación de enlaces entrantes y salientes.
2. Para contribuir positivamente al posicionamiento orgánico de un sitio web en los motores de búsqueda se puede:
 - a) Crear contenido relevante y de alta calidad que satisfaga las necesidades de información de los usuarios.
 - b) Obtener enlaces entrantes de otros sitios web de alta autoridad y relevancia temática.
 - c) Incluir una densidad alta de palabras clave para asegurar que el sitio web aparezca en tantas búsquedas como sea posible.
 - d) Mejorar la velocidad de carga del sitio web y asegurar que sea *responsive* y fácil de usar en dispositivos móviles.
3. La web actual se enfrenta a problemas como:
 - a) Presencia de una alta calidad en los datos.
 - b) Heterogeneidad en los datos.
 - c) Presencia de datos volátiles y distribuidos.
 - d) Presencia de grandes volúmenes de datos estructurados.
4. Dentro del ecosistema de Hadoop, el HDFS se caracteriza por:
 - a) La capacidad ilimitada de almacenamiento.
 - b) La tolerancia a fallos mediante la replicación de datos.
 - c) El almacenamiento exclusivo para archivos de texto.
 - d) El modelo de acceso y de escritura de datos en tiempo real.
5. En una empresa comercial, se tiene una red donde los nodos corresponden a los empleados y las aristas representan la frecuencia con la que colaboran en las ventas. Se busca mejorar la colaboración entre los empleados para aumentar las ventas totales, por lo que la directiva debe:
 - a) No tener en cuenta la centralidad de cercanía de los empleados al diseñar estrategias de colaboración, ya que esto podría introducir complicaciones adicionales en el proceso de trabajo.
 - b) Utilizar la centralidad de cercanía para identificar los empleados menos cercanos a otros en la red y asignarles tareas individuales para evitar posibles conflictos y desacuerdos en el proceso de colaboración.
 - c) No considerar la centralidad de cercanía de los empleados al diseñar estrategias de colaboración, ya que todos los empleados tienen el mismo acceso a la información y los recursos.
 - d) Utilizar la centralidad de cercanía para identificar los empleados más cercanos a otros en la red y promover la colaboración entre ellos, facilitando así la comunicación y el intercambio de conocimientos para mejorar las ventas.
6. ¿Cuál de las siguientes opciones describe mejor la diferencia clave entre Web Crawling y Web Scraping?
 - a) Web Scraping implica el análisis de la estructura y el contenido de las páginas web para extraer datos, mientras que Web Crawling se refiere a la descarga y almacenamiento de páginas web completas.
 - b) Web Scraping es más eficaz para rastrear e indexar contenido web para motores de búsqueda, mientras que Web Crawling se utiliza principalmente para la extracción de datos en proyectos de investigación.
 - c) Web Crawling se centra en la exploración y recopilación de enlaces de múltiples sitios web, mientras que Web Scraping se enfoca en la extracción específica de datos de páginas web individuales.

- d) Web Crawling se realiza utilizando herramientas de automatización como Selenium WebDriver, mientras que Web Scraping se lleva a cabo mediante el análisis de HTML y CSS.
7. Considerando las prácticas éticas y legales en el Web Scraping, se puede asegurar que:
- Web Scraping sobre datos personales sin consentimiento es generalmente aceptado si los datos se utilizan con fines de investigación.
 - Es importante revisar y respetar los términos de servicio del sitio web, así como las leyes aplicables de protección de datos y derechos de autor, antes de realizar Web Scraping.
 - Web Scraping requiere considerar las políticas de `robots.txt` del sitio web objetivo.
 - La extracción de datos mediante Web Scraping siempre es legal, independientemente de las leyes locales sobre derechos de autor y privacidad de datos definidos en los sitios web.
8. En un sistema de control de tráfico urbano basado en reglas, ¿cuál de las siguientes reglas sería más efectiva para manejar situaciones de congestión en una intersección?
- Si hay pocos vehículos en la intersección, reducir el tiempo de los semáforos en verde.
 - Si hay muchos vehículos en la intersección, aumentar el tiempo de los semáforos en verde.
 - Si hay un vehículo de emergencia en la intersección, detener todos los demás vehículos.
 - Si hay muchos peatones cruzando la intersección, reducir el tiempo de los semáforos en rojo.
9. Considerando las dimensiones y desafíos inherentes a Big Data puede afirmarse, tomando en cuenta las características clave y las implicaciones para su procesamiento y análisis, que:
- Big Data se caracteriza principalmente por su pequeño volumen y uniformidad, permitiendo un procesamiento eficiente con mínimas adaptaciones de las herramientas de análisis de datos tradicionales.
 - Big Data no desafía la capacidad de las herramientas tradicionales de procesamiento de datos para capturar, almacenar, gestionar y analizar efectivamente la información, dada la evolución constante de las capacidades computacionales y algoritmos de optimización.
 - Aunque Big Data puede incluir datos estructurados, su naturaleza se expande al incorporar grandes cantidades de datos no estructurados y semiestructurados, lo que exige el uso de tecnologías especializadas en almacenamiento y procesamiento como Hadoop y sistemas de bases de datos NoSQL.
10. La integración de Hadoop y MapReduce en la RI trae como ventaja:
- La posibilidad de la extracción de información relevante y la generación de resultados significativos de grandes conjuntos de datos.
 - La reducción de los costos operativos a cero.
 - La eliminación de la necesidad de sistemas de bases de datos.
 - La garantía de la privacidad absoluta de los datos procesados.
11. La Web 3 se conoce como:
- Web de solo lectura.
 - Internet de las cosas.
 - Web de escritura-lectura.
 - Web semántica.
12. En un SRI la indexación:
- Mejora la experiencia del usuario.
 - Consiste en asociar un identificador único a cada dato almacenado en el sistema.
 - Facilita la RI relevante.
 - Permite la organización y la categorización de la información.
13. En el contexto de la RI en redes sociales se puede afirmar que:
- El análisis de redes sociales se centra exclusivamente en la cantidad de seguidores que tiene un usuario en particular para determinar su influencia en la red.
 - El análisis de redes sociales no es útil para comprender la difusión de información en una red social específica, para ello se utiliza la medida de centralidad de vector propio.
 - El análisis de centralidad de intermediación se utiliza para identificar usuarios que son importantes en una red social debido a su posición como “puentes” entre diferentes grupos de usuarios.
 - El análisis de sentimientos se utiliza para determinar la popularidad de una publicación en redes sociales sin tener en cuenta la opinión de los usuarios.

14. En el algoritmo de Indexación Basada en Clasificación Bloqueada (BSBI), ¿cuál es el paso final para crear un índice invertido para la colección completa de los datos?
- a) Dividir la colección de datos en bloques de tamaño fijo.
 - b) Eliminar los términos duplicados de los índices.
 - c) Indexar cada bloque de forma independiente.
 - d) Fusionar los índices invertidos de cada bloque.
15. Se puede afirmar que:
- a) La cantidad de conexiones de un nodo siempre indica su influencia en la red.
 - b) Todas las relaciones en una red tienen la misma importancia para el análisis.
 - c) No es posible identificar subgrupos dentro de una red utilizando análisis de redes.
 - d) El tamaño de una red es siempre indicativo de su efectividad en la transmisión de información.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. La premisa básica del algoritmo de PageRank para clasificar páginas web en los resultados de búsqueda es:
 - a) Los enlaces entrantes a una página web desde otras páginas contribuyen a su importancia y clasificación.
 - b) La antigüedad de la página web es el principal factor para determinar su clasificación.
 - c) La cantidad de visitas que recibe una página web determina su clasificación.
 - d) El contenido y la relevancia de las palabras clave en la página web determinan su posición.
2. El posicionamiento de un sitio web en los motores de búsqueda puede ser afectado por:
 - a) Mantener una estructura de URL clara y coherente.
 - b) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - c) Contenido sin valor en el sitio web.
 - d) Obtener enlaces de sitios web irrelevantes y de baja calidad.
3. La diferencia entre la indexación por tokens y la indexación por conceptos puede definirse como:
 - a) La indexación por tokens normaliza los datos reduciéndolos a su forma básica, mientras que la indexación por conceptos utiliza un método de ordenación para organizar los términos característicos de los datos.
 - b) La indexación por tokens asigna un valor numérico a cada término de los datos, mientras que la indexación por conceptos utiliza algoritmos de encriptación para proteger la privacidad de los datos.
 - c) La indexación por tokens asigna pesos a los términos basados en su importancia relativa, mientras que la indexación por conceptos utiliza un sistema de etiquetado para asociar términos con características generales.
 - d) La indexación por tokens divide los datos en términos individuales, mientras que la indexación por conceptos agrupa los datos en categorías definidas.
4. La política de ordenación de URLs en los Web Crawlers tiene como aspecto fundamental:
 - a) Limitar el acceso de los crawlers a ciertas secciones de un sitio web, evitando el rastreo de URLs consideradas menos importantes o sensibles.
 - b) Establecer la prioridad de rastreo de las URLs, determinando el orden en que los crawlers visitan y procesan cada página web.
 - c) Definir la estructura de la URL de destino, asegurando que estén ordenadas alfabéticamente para facilitar la navegación y la indexación.
 - d) Determinar la forma en que los crawlers asignan un valor de relevancia a cada URL para clasificarlas en el índice de búsqueda.
5. ¿Qué algoritmos permiten obtener información de una red?
 - a) Índices de centralidad.
 - b) Detección de comunidades.
 - c) Hypertext Induced Topic Selection (HITS).
 - d) PageRank.
6. En el algoritmo de Indexación Basada en Clasificación Bloqueada (BSBI), ¿cuál es el paso final para crear un índice invertido para la colección completa de los datos?
 - a) Dividir la colección de datos en bloques de tamaño fijo.
 - b) Eliminar los términos duplicados de los índices.
 - c) Indexar cada bloque de forma independiente.
 - d) Fusionar los índices invertidos de cada bloque.
7. El componente responsable de la gestión de recursos y planificación de tareas en Hadoop es:
 - a) MapReduce.
 - b) HDFS.
 - c) Hadoop Common.
 - d) YARN.
8. En la RI en el contexto de Big Data, se puede asegurar que:
 - a) El almacenamiento distribuido es una técnica obsoleta.

- b) El almacenamiento distribuido centraliza todos los datos en un único servidor para facilitar su gestión y mantenimiento.
 - c) El almacenamiento distribuido reparte los datos en múltiples servidores para mejorar la disponibilidad y la redundancia del sistema.
 - d) El almacenamiento distribuido ofrece ventajas significativas en términos de escalabilidad y rendimiento en comparación con el almacenamiento centralizado.
9. En el contexto de la representación del conocimiento basada en herencia, ¿qué caracteriza a la herencia cancelable?
- a) Las propiedades heredadas siempre se mantienen y no pueden anularse.
 - b) La herencia es el resultado del razonamiento no transitivo.
 - c) Las conclusiones no están determinadas y dependen del nodo de interés.
 - d) No existe ambigüedad en las conclusiones obtenidas.
10. Dentro del análisis de redes, la centralidad de grado mide la importancia de un nodo basándose en:
- a) La cantidad de vecinos del nodo.
 - b) El número de aristas que posee el nodo.
 - c) La cantidad de veces que aparece el nodo en el camino mínimo entre cualquier par de nodos.
 - d) El grado del nodo.
11. Sobre WordNet y su aplicación en el procesamiento del lenguaje natural, se puede afirmar que:
- a) Aunque WordNet es una herramienta valiosa en el NLP, su estructura no incluye información sobre la frecuencia de uso de las palabras en el lenguaje natural.
 - b) El diseño de WordNet facilita su integración en aplicaciones multilingües de NLP, aunque su desarrollo original se centró en el inglés.
 - c) Los synsets facilitan la identificación de relaciones semánticas entre palabras, como la hiperonimia y la meronimia, enriqueciendo tareas de NLP.
 - d) WordNet proporciona una base para la desambiguación semántica al agrupar palabras con significados similares en synsets.
12. El algoritmo de PageRank converge si:
- a) Se define un factor de normalización en la fórmula de la función.
 - b) La norma de la diferencia entre los vectores es menor a un umbral predefinido.
 - c) Finaliza la ejecución cuando el número de iteraciones excede un máximo de iteraciones previamente definido.
 - d) El algoritmo no se implementa de forma iterativa.
13. La Web 2.0 se caracteriza por:
- a) Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
 - b) Las plataformas ofrecen experiencias a la medida, permitiendo a los usuarios personalizar sus perfiles, recibir recomendaciones ajustadas al contenido y participar en filtrado colaborativo.
 - c) La existencia de sitios web dinámicos e interactivos que permiten a los usuarios participar, comentar e interactuar tanto con los creadores de contenido como con otros usuarios.
 - d) El uso de distintas tecnologías para crear experiencias web más interactivas y con mayor capacidad de respuesta.
14. ¿Qué estrategia utilizan los Web Crawlers para asegurar un rastreo eficiente y respetuoso de los recursos de los sitios web?
- a) Extraer únicamente contenido multimedia para reducir la carga en los servidores web.
 - b) Visitar y rastrear todos los enlaces de una página web simultáneamente.
 - c) Ignorar completamente el archivo `robots.txt` de los sitios web.
 - d) Seguir las directrices del archivo `robots.txt` y aplicar un retraso entre las solicitudes.
15. Se puede afirmar que:
- a) Uno de los desafíos en el procesamiento de Big Data es la capacidad de gestionar y analizar datos provenientes de diversas fuentes y en diferentes formatos de manera eficiente.
 - b) MapReduce es un modelo de procesamiento distribuido utilizado para trabajar con grandes volúmenes de datos.
 - c) Los SRI pueden beneficiarse de MapReduce para mejorar la RI relevante.
 - d) El término “Big Data” se refiere exclusivamente al volumen de datos que una organización maneja, sin tener en cuenta la velocidad, la variedad y la veracidad de los datos.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Dentro del ecosistema de Hadoop, el HDFS se caracteriza por:
 - a) La capacidad ilimitada de almacenamiento.
 - b) La tolerancia a fallos mediante la replicación de datos.
 - c) El almacenamiento exclusivo para archivos de texto.
 - d) El modelo de acceso y de escritura de datos en tiempo real.
2. Considerando las dimensiones y desafíos inherentes a Big Data puede afirmarse, tomando en cuenta las características clave y las implicaciones para su procesamiento y análisis, que:
 - a) Big Data se caracteriza principalmente por su pequeño volumen y uniformidad, permitiendo un procesamiento eficiente con mínimas adaptaciones de las herramientas de análisis de datos tradicionales.
 - b) Big Data no desafía la capacidad de las herramientas tradicionales de procesamiento de datos para capturar, almacenar, gestionar y analizar efectivamente la información, dada la evolución constante de las capacidades computacionales y algoritmos de optimización.
 - c) Aunque Big Data puede incluir datos estructurados, su naturaleza se expande al incorporar grandes cantidades de datos no estructurados y semiestructurados, lo que exige el uso de tecnologías especializadas en almacenamiento y procesamiento como Hadoop y sistemas de bases de datos NoSQL.
 - d) Además de su complejidad y diversidad, Big Data introduce desafíos significativos en términos de veracidad y variabilidad al requerir métodos avanzados de limpieza y validación de datos para asegurar la integridad del análisis.
3. La centralidad de intermediación de un nodo indica:
 - a) La frecuencia con la que un nodo actúa como puente en el camino más corto entre otros dos nodos.
 - b) El número total de conexiones entrantes y salientes.
 - c) La resistencia del nodo a fallos.
 - d) La cantidad de nodos vecinos directos.
4. En el algoritmo de Indexación Basada en Clasificación Bloqueada (BSBI), ¿cuál es el paso final para crear un índice invertido para la colección completa de los datos?
 - a) Dividir la colección de datos en bloques de tamaño fijo.
 - b) Eliminar los términos duplicados de los índices.
 - c) Indexar cada bloque de forma independiente.
 - d) Fusionar los índices invertidos de cada bloque.
5. Si una red cumple la propiedad de ser un grafo de mundo pequeño, entonces se conoce que:
 - a) La red posee pocos vértices.
 - b) La longitud media del camino entre todo par de vértices es pequeña.
 - c) El número de componentes fuertemente conexas está relacionado con la cantidad de grafos K_n presentes.
 - d) El grafo es un anillo regular de grado 5.
6. Se tiene un grafo donde cada nodo es un personaje de cierto libro de cuentos y la existencia de las aristas está definida si dos personajes aparecen en el mismo cuento. Se puede asegurar que:
 - a) La centralidad de grado es útil para identificar los personajes con más conexiones dentro de la red, lo que puede indicar su importancia en el libro.
 - b) La centralidad de cercanía indica el grado de conexión de cada personaje con el resto de los personajes de los cuentos del libro.
 - c) La centralidad de intermediación identifica a los personajes que actúan como conectores entre personajes de cuentos distintos.
 - d) La centralidad de cercanía ofrece una relación entre la cantidad de vecinos de un nodo con respecto a la longitud máxima de un camino dentro del grafo partiendo del nodo en cuestión.
7. Al “relajar” el concepto de clique en la detección de comunidades se intenta solucionar:
 - a) La uniformidad de los nodos en términos de grado.

- b) La necesidad de datos externos para analizar la red.
 - c) La imposibilidad de conectar nodos distantes.
 - d) El solapamiento y la complejidad computacional.
8. En un sistema de control de tráfico urbano basado en reglas, ¿cuál de las siguientes reglas sería más efectiva para manejar situaciones de congestión en una intersección?
- a) Si hay pocos vehículos en la intersección, reducir el tiempo de los semáforos en verde.
 - b) Si hay muchos vehículos en la intersección, aumentar el tiempo de los semáforos en verde.
 - c) Si hay un vehículo de emergencia en la intersección, detener todos los demás vehículos.
 - d) Si hay muchos peatones cruzando la intersección, reducir el tiempo de los semáforos en rojo.
9. En una biblioteca digital se necesita implementar un sistema que permita a los usuarios encontrar libros y artículos científicos de forma rápida y precisa. Los documentos están en diversos formatos, incluyendo PDF, EPUB y HTML. Se requiere seleccionar un algoritmo de indexación adecuado para el sistema, por lo que el programador designado para la implementación debe considerar:
- a) La capacidad para manejar documentos en diferentes formatos.
 - b) La velocidad de indexación y recuperación de datos.
 - c) La complejidad del algoritmo en términos de implementación y mantenimiento.
 - d) La capacidad del algoritmo para procesar imágenes incrustadas o referenciadas en los ficheros.
10. ¿Qué estrategia utilizan los Web Crawlers para asegurar un rastreo eficiente y respetuoso de los recursos de los sitios web?
- a) Extraer únicamente contenido multimedia para reducir la carga en los servidores web.
 - b) Visitar y rastrear todos los enlaces de una página web simultáneamente.
 - c) Ignorar completamente el archivo `robots.txt` de los sitios web.
 - d) Seguir las directrices del archivo `robots.txt` y aplicar un retraso entre las solicitudes.
11. Un SRI es capaz de:
- a) No necesitar de ningún almacenamiento externo para alojar los índices de los datos.
- b) Reducir el tiempo de indexación de los datos si utiliza servidores distribuidos dentro de la red para que cada uno ejecute la indexación del mismo conjunto de datos.
 - c) Generar índices invertidos de manera óptima sin considerar el contexto.
 - d) Crear los índices asociados a los datos sin tener que analizar cada dato.
12. Al realizar la optimización de contenido para SEO debe considerarse:
- a) Utilizar etiquetas de título y meta descripciones únicas y relevantes para cada página.
 - b) Seleccionar las palabras clave al azar.
 - c) Crear contenido valioso y original que satisfaga las necesidades de los usuarios.
 - d) Incluir palabras clave de manera excesiva en el contenido para mejorar el posicionamiento.
13. Un investigador necesita recopilar datos de múltiples sitios web para un estudio académico, pero se enfrenta a varios desafíos al realizar el proceso de Web Scraping de manera ética y legal. ¿Cuál de las siguientes opciones describe mejor uno de los desafíos asociados al proceso de Web Scraping?
- a) La dificultad para encontrar herramientas de Web Scraping gratuitas y fiables.
 - b) La disponibilidad limitada de datos en línea que se pueden extraer utilizando técnicas de Web Scraping.
 - c) La necesidad de estar montado sobre un crawler que cumpla con todas las políticas.
 - d) La necesidad de comprender la estructura del sitio web y su código HTML para extraer los datos correctamente.
14. Para contribuir positivamente al posicionamiento orgánico de un sitio web en los motores de búsqueda se puede:
- a) Crear contenido relevante y de alta calidad que satisfaga las necesidades de información de los usuarios.
 - b) Obtener enlaces entrantes de otros sitios web de alta autoridad y relevancia temática.
 - c) Incluir una densidad alta de palabras clave para asegurar que el sitio web aparezca en tantas búsquedas como sea posible.
 - d) Mejorar la velocidad de carga del sitio web y asegurar que sea *responsive* y fácil de usar en dispositivos móviles.
15. En el contexto del procesamiento de grandes conjuntos de datos, se puede asegurar que:
- a) MapReduce no es adecuado para el procesamiento de datos no estructurados.

- b) MapReduce solo puede manejar pequeñas cantidades de datos y no escala bien a grandes conjuntos de datos.
- c) MapReduce es un enfoque para procesar datos de forma secuencial en un solo servidor para evitar problemas de concurrencia.
- d) MapReduce divide una tarea en múltiples pasos de map y reduce que se ejecutan de forma secuencial en diferentes servidores para mejorar el rendimiento y la escalabilidad.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. En el diseño de un videojuego de roles (RPG) en el que los personajes tienen atributos como salud, fuerza y velocidad, ¿cuál de las siguientes opciones representa mejor una implementación de la representación del conocimiento orientado a objetos?
 - a) Cada personaje se representa como un objeto con propiedades como salud, fuerza y velocidad, y métodos para modificar y consultar estos valores.
 - b) Cada personaje se representa como una función que calcula sus atributos en función de su nivel y experiencia.
 - c) Cada personaje se representa como una matriz de números que almacena sus valores de atributos.
 - d) Cada personaje se representa como una lista de cadenas de texto que describen sus características físicas y habilidades.
2. La premisa básica del algoritmo de PageRank para clasificar páginas web en los resultados de búsqueda es:
 - a) Los enlaces entrantes a una página web desde otras páginas contribuyen a su importancia y clasificación.
 - b) La antigüedad de la página web es el principal factor para determinar su clasificación.
 - c) La cantidad de visitas que recibe una página web determina su clasificación.
 - d) El contenido y la relevancia de las palabras clave en la página web determinan su posición.
3. El algoritmo de PageRank converge si:
 - a) Se define un factor de normalización en la fórmula de la función.
 - b) La norma de la diferencia entre los vectores es menor a un umbral predefinido.
 - c) Finaliza la ejecución cuando el número de iteraciones excede un máximo de iteraciones previamente definido.
 - d) El algoritmo no se implementa de forma iterativa.
4. En el contexto de la RI en redes, se puede afirmar que:
 - a) La detección de comunidades ayuda a identificar grupos de nodos altamente conectados entre sí, lo que puede ser útil para comprender la estructura y el contenido de la red.
 - b) La detección de comunidades en una red siempre produce resultados objetivos y consistentes independientemente del algoritmo utilizado.
 - c) La detección de comunidades solo se aplica a redes pequeñas y simples, no a redes grandes y complejas.
 - d) La detección de comunidades no es relevante para la RI en redes, ya que se centra únicamente en la estructura de la red sin considerar el contenido.
5. ¿Qué algoritmos permiten obtener información de una red?
 - a) Índices de centralidad.
 - b) Detección de comunidades.
 - c) Hypertext Induced Topic Selection (HITS).
 - d) PageRank.
6. Si una red cumple la propiedad de ser un grafo de mundo pequeño, entonces se conoce que:
 - a) La red posee pocos vértices.
 - b) La longitud media del camino entre todo par de vértices es pequeña.
 - c) El número de componentes fuertemente conexas está relacionado con la cantidad de grafos K_n presentes.
 - d) El grafo es un anillo regular de grado 5.
7. En la RI en el contexto de Big Data, se puede asegurar que:
 - a) El almacenamiento distribuido es una técnica obsoleta.
 - b) El almacenamiento distribuido centraliza todos los datos en un único servidor para facilitar su gestión y mantenimiento.
 - c) El almacenamiento distribuido reparte los datos en múltiples servidores para mejorar la disponibilidad y la redundancia del sistema.
 - d) El almacenamiento distribuido ofrece ventajas significativas en términos de escalabilidad y rendimiento en comparación con el almacenamiento centralizado.

8. El concepto de *rank sink* en el algoritmo de PageRank representa páginas web:
 - a) Con una baja calidad de contenido y una cantidad insuficiente de enlaces salientes, lo que las hace menos relevantes en los resultados de búsqueda.
 - b) Con distribuciones uniformes de la puntuación de PageRank.
 - c) Con un alto número de enlaces salientes que no reciben enlaces entrantes, lo que puede afectar negativamente su puntuación de PageRank.
 - d) Que tienen una puntuación de PageRank más alta que otras debido a la manipulación de enlaces entrantes y salientes.
9. Referente al Web Crawler puede afirmarse que:
 - a) Los hipervínculos encontrados en cada sitio web que no pertenecen al dominio donde fueron encontrados se desechan, puesto que no expande el conjunto de URLs sin visitar.
 - b) Las páginas visitadas no se procesan nunca más.
 - c) No necesita de un conjunto inicial de URLs para recorrer la Web.
 - d) No tiene como objetivo indexar y recopilar información de diferentes sitios web.
10. Un SRI es capaz de:
 - a) No necesitar de ningún almacenamiento externo para alojar los índices de los datos.
 - b) Reducir el tiempo de indexación de los datos si utiliza servidores distribuidos dentro de la red para que cada uno ejecute la indexación del mismo conjunto de datos.
 - c) Generar índices invertidos de manera óptima sin considerar el contexto.
 - d) Crear los índices asociados a los datos sin tener que analizar cada dato.
11. En una red de transporte donde cada nodo es una parada de autobús y las aristas representan si existe un carro que pasa por ambos sitios, ¿qué puede mejorar el sistema de transporte?
 - a) Utilizar la centralidad de intermediación para identificar las paradas de transferencia clave y establecer nuevas conexiones entre diferentes líneas de transporte público.
 - b) Utilizar el grafo inducido de los nodos con mayor valor en la centralidad de grado para aplicar la centralidad de intermediación con el propósito de reforzar las paradas con mayor tráfico.
 - c) No implementar sistemas de información en tiempo real para los usuarios, ya que pueden aumentar la carga de los trabajadores.
 - d) No considerar la centralidad de intermediación de las estaciones de transporte público al planificar rutas y horarios, ya que no tiene impacto en las conexiones entre las paradas.
12. La transición de la Web 1.0 a la Web 2.0 se caracterizó principalmente por:
 - a) La reducción en el uso de HTML y CSS en el desarrollo de sitios web.
 - b) La disminución de la importancia de los motores de búsqueda en la navegación web.
 - c) El cambio de páginas web estáticas a dinámicas, permitiendo la interacción del usuario y la generación de contenido.
 - d) El aumento en la velocidad de conexión a internet, que permitió una mejor calidad de las páginas web.
13. En el contexto de la representación del conocimiento basada en herencia, ¿qué caracteriza a la herencia cancelable?
 - a) Las propiedades heredadas siempre se mantienen y no pueden anularse.
 - b) La herencia es el resultado del razonamiento no transitivo.
 - c) Las conclusiones no están determinadas y dependen del nodo de interés.
 - d) No existe ambigüedad en las conclusiones obtenidas.
14. El componente responsable de la gestión de recursos y planificación de tareas en Hadoop es:
 - a) MapReduce.
 - b) HDFS.
 - c) Hadoop Common.
 - d) YARN.
15. El algoritmo de PageRank puede describirse como un procedimiento utilizado para:
 - a) Determinar la velocidad de carga de una página web en un navegador.
 - b) Calcular la relevancia de una página web en función de la cantidad y calidad de los enlaces que apuntan hacia ella.
 - c) Clasificar las páginas web en función de su edad y autoridad.
 - d) Prevenir el spam y el contenido no deseado en las páginas web.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. En una biblioteca digital se necesita implementar un sistema que permita a los usuarios encontrar libros y artículos científicos de forma rápida y precisa. Los documentos están en diversos formatos, incluyendo PDF, EPUB y HTML. Se requiere seleccionar un algoritmo de indexación adecuado para el sistema, por lo que el programador designado para la implementación debe considerar:
 - a) La capacidad para manejar documentos en diferentes formatos.
 - b) La velocidad de indexación y recuperación de datos.
 - c) La complejidad del algoritmo en términos de implementación y mantenimiento.
 - d) La capacidad del algoritmo para procesar imágenes incrustadas o referenciadas en los ficheros.
2. En la RI en el contexto de Big Data, se puede asegurar que:
 - a) La indexación distribuida es una técnica obsoleta.
 - b) La indexación distribuida almacena todos los datos en un solo servidor para facilitar su acceso y búsqueda.
 - c) La indexación distribuida divide los datos en múltiples fragmentos que se almacenan en varios nodos para permitir búsquedas paralelas y mejorar la escalabilidad.
 - d) La indexación distribuida no ofrece ventajas en términos de rendimiento y escalabilidad en comparación con la indexación centralizada.
3. El componente responsable de la gestión de recursos y planificación de tareas en Hadoop es:
 - a) MapReduce.
 - b) HDFS.
 - c) Hadoop Common.
 - d) YARN.
4. Existen varias estrategias consideradas como no recomendables para mejorar el SEO de un sitio web. Dentro de estas estrategias negativas, se encuentran:
 - a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Evitar el uso de las meta etiquetas y descripciones del sitio web para reflejar el contenido de la página.
 - c) Construir enlaces naturales de sitios web con autoridad y relevancia temática.
 - d) Incluir una cantidad excesiva de palabras clave irrelevantes para intentar manipular los rankings de búsqueda.
5. En el contexto de la RI en redes sociales se puede afirmar que:
 - a) El análisis de redes sociales se centra exclusivamente en la cantidad de seguidores que tiene un usuario en particular para determinar su influencia en la red.
 - b) El análisis de redes sociales no es útil para comprender la difusión de información en una red social específica, para ello se utiliza la medida de centralidad de vector propio.
 - c) El análisis de centralidad de intermediación se utiliza para identificar usuarios que son importantes en una red social debido a su posición como “puentes” entre diferentes grupos de usuarios.
 - d) El análisis de sentimientos se utiliza para determinar la popularidad de una publicación en redes sociales sin tener en cuenta la opinión de los usuarios.
6. Los algoritmos para detectar comunidades en una red intentan:
 - a) Buscar subgrafos tal que no incluyan nodos cuya ausencia desconecte al subgrafo.
 - b) Seleccionar aleatoriamente nodos de alto grado y sus vecinos.
 - c) Buscar conjuntos donde cada nodo de un mismo conjunto tenga características similares al resto de los nodos del conjunto.
 - d) Encontrar grupos donde los nodos pertenecientes a los mismos grupos son cercanos bajo cierta métrica y lejanos con respecto a los nodos de otros grupos.
7. La Web 2.0 se caracteriza por:

- a) Las actualizaciones del contenido de los sitios web se efectúa de forma manual dentro del código HTML
 - b) Las plataformas ofrecen experiencias a la medida, permitiendo a los usuarios personalizar sus perfiles, recibir recomendaciones ajustadas al contenido y participar en filtrado colaborativo.
 - c) La existencia de sitios web dinámicos e interactivos que permiten a los usuarios participar, comentar e interactuar tanto con los creadores de contenido como con otros usuarios.
 - d) El uso de distintas tecnologías para crear experiencias web más interactivas y con mayor capacidad de respuesta.
8. La premisa básica del algoritmo de PageRank para clasificar páginas web en los resultados de búsqueda es:
- a) Los enlaces entrantes a una página web desde otras páginas contribuyen a su importancia y clasificación.
 - b) La antigüedad de la página web es el principal factor para determinar su clasificación.
 - c) La cantidad de visitas que recibe una página web determina su clasificación.
 - d) El contenido y la relevancia de las palabras clave en la página web determinan su posición.
9. Referente al Web Crawler puede afirmarse que:
- a) Los hipervínculos encontrados en cada sitio web que no pertenecen al dominio donde fueron encontrados se desechan, puesto que no expande el conjunto de URLs sin visitar.
 - b) Las páginas visitadas no se procesan nunca más.
 - c) No necesita de un conjunto inicial de URLs para recorrer la Web.
 - d) No tiene como objetivo indexar y recopilar información de diferentes sitios web.
10. En un SRI la indexación:
- a) Mejora la experiencia del usuario.
 - b) Consiste en asociar un identificador único a cada dato almacenado en el sistema.
 - c) Facilita la RI relevante.
 - d) Permite la organización y la categorización de la información.
11. En una red de transporte donde cada nodo es una parada de autobús y las aristas representan si existe un carro que pasa por ambos sitios, ¿qué puede mejorar el sistema de transporte?
- a) Utilizar la centralidad de intermediación para identificar las paradas de transferencia clave y establecer nuevas conexiones entre diferentes líneas de transporte público.
 - b) Utilizar el grafo inducido de los nodos con mayor valor en la centralidad de grado para aplicar la centralidad de intermediación con el propósito de reforzar las paradas con mayor tráfico.
 - c) No implementar sistemas de información en tiempo real para los usuarios, ya que pueden aumentar la carga de los trabajadores.
 - d) No considerar la centralidad de intermediación de las estaciones de transporte público al planificar rutas y horarios, ya que no tiene impacto en las conexiones entre las paradas.
12. Se puede afirmar que:
- a) La cantidad de conexiones de un nodo siempre indica su influencia en la red.
 - b) Todas las relaciones en una red tienen la misma importancia para el análisis.
 - c) No es posible identificar subgrupos dentro de una red utilizando análisis de redes.
 - d) El tamaño de una red es siempre indicativo de su efectividad en la transmisión de información.
13. En un sistema de control de tráfico urbano basado en reglas, ¿cuál de las siguientes reglas sería más efectiva para manejar situaciones de congestión en una intersección?
- a) Si hay pocos vehículos en la intersección, reducir el tiempo de los semáforos en verde.
 - b) Si hay muchos vehículos en la intersección, aumentar el tiempo de los semáforos en verde.
 - c) Si hay un vehículo de emergencia en la intersección, detener todos los demás vehículos.
 - d) Si hay muchos peatones cruzando la intersección, reducir el tiempo de los semáforos en rojo.
14. Con respecto a WordNet y su contribución a las ontologías y la representación del conocimiento en los sistemas de RI, puede afirmarse que:
- a) En WordNet, una palabra está asociada a un conjunto de sinónimos (synsets), siendo estas palabras intercambiables en un contexto.
 - b) WordNet diferencia claramente entre los significados de palabras según su uso en diferentes contextos, lo que permite aplicaciones avanzadas en desambiguación semántica más allá de los sistemas de recomendación y RI.
 - c) La integración de WordNet en sistemas de RI permite la expansión de consultas y la mejora de la precisión de los resultados al entender mejor el

significado de los términos de búsqueda a través de su contexto semántico y las relaciones entre palabras.

- d) Una limitación de WordNet en la representación del conocimiento es su enfoque en el idioma inglés, lo que plantea desafíos en la aplicación global y la interoperabilidad con sistemas de información multilingües.

15. La centralidad de intermediación de un nodo in-

dica:

- a) La frecuencia con la que un nodo actúa como puente en el camino más corto entre otros dos nodos.
- b) El número total de conexiones entrantes y salientes.
- c) La resistencia del nodo a fallos.
- d) La cantidad de nodos vecinos directos.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Dentro del análisis de redes, la centralidad de grado mide la importancia de un nodo basándose en:
 - a) La cantidad de vecinos del nodo.
 - b) El número de aristas que posee el nodo.
 - c) La cantidad de veces que aparece el nodo en el camino mínimo entre cualquier par de nodos.
 - d) El grado del nodo.
2. La web actual se enfrenta a problemas como:
 - a) Presencia de una alta calidad en los datos.
 - b) Heterogeneidad en los datos.
 - c) Presencia de datos volátiles y distribuidos.
 - d) Presencia de grandes volúmenes de datos estructurados.
3. Una buena práctica de SEO para mejorar el posicionamiento de un sitio web en los motores de búsqueda es:
 - a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Copiar contenido directamente de otros sitios web populares para aumentar la cantidad de páginas indexadas.
 - c) Llenar el contenido con palabras clave irrelevantes para aumentar la densidad de palabras clave.
 - d) Obtener enlaces de otros sitios web relevantes y de calidad que apunten al sitio.
4. En la RI en el contexto de Big Data, se puede asegurar que:
 - a) La indexación distribuida es una técnica obsoleta.
 - b) La indexación distribuida almacena todos los datos en un solo servidor para facilitar su acceso y búsqueda.
 - c) La indexación distribuida divide los datos en múltiples fragmentos que se almacenan en varios nodos para permitir búsquedas paralelas y mejorar la escalabilidad.
 - d) La indexación distribuida no ofrece ventajas en términos de rendimiento y escalabilidad en comparación con la indexación centralizada.
5. Analizar una red permite:
 - a) Evaluar la calidad del contenido de un sitio web.
 - b) Obtener predicciones exactas de eventos futuros en mercados financieros.
 - c) Detectar posibles tendencias antes de que se conviertan en tendencia.
 - d) Encontrar nodos “sensibles” o críticos para la red.
6. En el contexto del procesamiento de grandes conjuntos de datos, se puede asegurar que:
 - a) MapReduce no es adecuado para el procesamiento de datos no estructurados.
 - b) MapReduce solo puede manejar pequeñas cantidades de datos y no escala bien a grandes conjuntos de datos.
 - c) MapReduce es un enfoque para procesar datos de forma secuencial en un solo servidor para evitar problemas de concurrencia.
 - d) MapReduce divide una tarea en múltiples pasos de map y reduce que se ejecutan de forma secuencial en diferentes servidores para mejorar el rendimiento y la escalabilidad.
7. En el contexto de la RI en redes sociales se puede afirmar que:
 - a) El análisis de redes sociales se centra exclusivamente en la cantidad de seguidores que tiene un usuario en particular para determinar su influencia en la red.
 - b) El análisis de redes sociales no es útil para comprender la difusión de información en una red social específica, para ello se utiliza la medida de centralidad de vector propio.
 - c) El análisis de centralidad de intermediación se utiliza para identificar usuarios que son importantes en una red social debido a su posición como “puentes” entre diferentes grupos de usuarios.
 - d) El análisis de sentimientos se utiliza para determinar la popularidad de una publicación en redes sociales sin tener en cuenta la opinión de los usuarios.

8. Considerando las dimensiones y desafíos inherentes a Big Data puede afirmarse, tomando en cuenta las características clave y las implicaciones para su procesamiento y análisis, que:
 - a) Big Data se caracteriza principalmente por su pequeño volumen y uniformidad, permitiendo un procesamiento eficiente con mínimas adaptaciones de las herramientas de análisis de datos tradicionales.
 - b) Big Data no desafía la capacidad de las herramientas tradicionales de procesamiento de datos para capturar, almacenar, gestionar y analizar efectivamente la información, dada la evolución constante de las capacidades computacionales y algoritmos de optimización.
 - c) Aunque Big Data puede incluir datos estructurados, su naturaleza se expande al incorporar grandes cantidades de datos no estructurados y semiestructurados, lo que exige el uso de tecnologías especializadas en almacenamiento y procesamiento como Hadoop y sistemas de bases de datos NoSQL.
 - d) Además de su complejidad y diversidad, Big Data introduce desafíos significativos en términos de veracidad y variabilidad al requerir métodos avanzados de limpieza y validación de datos para asegurar la integridad del análisis.
9. En el diseño de un videojuego de roles (RPG) en el que los personajes tienen atributos como salud, fuerza y velocidad, ¿cuál de las siguientes opciones representa mejor una implementación de la representación del conocimiento orientado a objetos?
 - a) Cada personaje se representa como un objeto con propiedades como salud, fuerza y velocidad, y métodos para modificar y consultar estos valores.
 - b) Cada personaje se representa como una función que calcula sus atributos en función de su nivel y experiencia.
 - c) Cada personaje se representa como una matriz de números que almacena sus valores de atributos.
 - d) Cada personaje se representa como una lista de cadenas de texto que describen sus características físicas y habilidades.
10. Un SRI es capaz de:
 - a) No necesitar de ningún almacenamiento externo para alojar los índices de los datos.
 - b) Reducir el tiempo de indexación de los datos si utiliza servidores distribuidos dentro de la red para que cada uno ejecute la indexación del mismo conjunto de datos.
 - c) Generar índices invertidos de manera óptima sin considerar el contexto.
 - d) Crear los índices asociados a los datos sin tener que analizar cada dato.
11. Sobre el algoritmo de PageRank, visto en clase, se puede afirmar que:
 - a) Evalúa la importancia de un sitio web en función de la calidad y cantidad de enlaces entrantes que recibe de otros sitios web.
 - b) Asigna una puntuación baja a los sitios web que contienen muchos enlaces salientes, ya que indica una falta de relevancia.
 - c) Asigna una puntuación alta a los sitios web que tienen un gran número de enlaces entrantes sin tener en cuenta la calidad de esos enlaces.
 - d) Solo tiene en cuenta el contenido en un sitio web para determinar su relevancia en los resultados de búsqueda.
12. El algoritmo de PageRank puede describirse como un procedimiento utilizado para:
 - a) Determinar la velocidad de carga de una página web en un navegador.
 - b) Calcular la relevancia de una página web en función de la cantidad y calidad de los enlaces que apuntan hacia ella.
 - c) Clasificar las páginas web en función de su edad y autoridad.
 - d) Prevenir el spam y el contenido no deseado en las páginas web.
13. La afirmación que mejor describe la política de amabilidad en los Web Crawlers es:
 - a) Los Web Crawlers son libres de recopilar datos de cualquier sitio web sin restricciones, independientemente de la cantidad de tráfico que generen.
 - b) La política de amabilidad establece pautas y reglas sobre cómo los crawlers deben interactuar con los sitios web para minimizar la carga del servidor y respetar las directivas de los administradores del sitio.
 - c) La política de amabilidad de los Web Crawlers dicta que los crawlers deben priorizar ciertos tipos de contenido sobre otros, ignorando completamente ciertas páginas web.
 - d) Los crawlers se diseñan para acceder a sitios web sin restricciones y extraer datos de manera agresiva para su indexación.
14. ¿Qué es un 'Uniform Resource Locator (URL) Frontier' en el contexto de Web Crawling?
 - a) Una base de datos que almacena URLs únicas identificadas como recursos en la web.

- b) Una técnica para filtrar URLs irrelevantes y mejorar la eficiencia del crawling.
- c) Un protocolo que define cómo se deben formatear las URLs para el crawling.
- d) Una lista prioritaria de URLs que aún no han sido visitadas por el crawler.
15. No se considera como técnica para detectar comunidades en una red:
- a) Analizar la mutualidad de los enlaces.
- b) Usar el agrupamiento jerárquico.
- c) Utilizar el algoritmo de K-Means.
- d) Encontrar cliques de vértices de grado par.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. En el modelo de representación del conocimiento basado en herencia se puede asegurar que:
 - a) La cantidad de padres que puede tener un nodo no es mayor que 4.
 - b) Solo se usa en entornos referentes a la biología.
 - c) El razonamiento deducido está respaldado por al menos un camino dentro del grafo.
 - d) Las conclusiones pueden ser canceladas si el grafo es ambiguo.
2. En la RI en el contexto de Big Data, se puede asegurar que:
 - a) El almacenamiento distribuido es una técnica obsoleta.
 - b) El almacenamiento distribuido centraliza todos los datos en un único servidor para facilitar su gestión y mantenimiento.
 - c) El almacenamiento distribuido reparte los datos en múltiples servidores para mejorar la disponibilidad y la redundancia del sistema.
 - d) El almacenamiento distribuido ofrece ventajas significativas en términos de escalabilidad y rendimiento en comparación con el almacenamiento centralizado.
3. En un SRI la indexación:
 - a) Mejora la experiencia del usuario.
 - b) Consiste en asociar un identificador único a cada dato almacenado en el sistema.
 - c) Facilita la RI relevante.
 - d) Permite la organización y la categorización de la información.
4. Si una red cumple la propiedad de ser un grafo de mundo pequeño, entonces se conoce que:
 - a) La red posee pocos vértices.
 - b) La longitud media del camino entre todo par de vértices es pequeña.
 - c) El número de componentes fuertemente conexas está relacionado con la cantidad de grafos K_n presentes.
 - d) El grafo es un anillo regular de grado 5.
5. En el diseño de un videojuego de roles (RPG) en el que los personajes tienen atributos como salud, fuerza y velocidad, ¿cuál de las siguientes opciones representa mejor una implementación de la representación del conocimiento orientado a objetos?
 - a) Cada personaje se representa como un objeto con propiedades como salud, fuerza y velocidad, y métodos para modificar y consultar estos valores.
 - b) Cada personaje se representa como una función que calcula sus atributos en función de su nivel y experiencia.
 - c) Cada personaje se representa como una matriz de números que almacena sus valores de atributos.
 - d) Cada personaje se representa como una lista de cadenas de texto que describen sus características físicas y habilidades.
6. En una biblioteca digital se necesita implementar un sistema que permita a los usuarios encontrar libros y artículos científicos de forma rápida y precisa. Los documentos están en diversos formatos, incluyendo PDF, EPUB y HTML. Se requiere seleccionar un algoritmo de indexación adecuado para el sistema, por lo que el programador designado para la implementación debe considerar:
 - a) La capacidad para manejar documentos en diferentes formatos.
 - b) La velocidad de indexación y recuperación de datos.
 - c) La complejidad del algoritmo en términos de implementación y mantenimiento.
 - d) La capacidad del algoritmo para procesar imágenes incrustadas o referenciadas en los ficheros.
7. Para la RI, el análisis de las redes puede:
 - a) Ayudar a identificar grupos de interés.
 - b) Indicar la importancia de una entidad en la transmisión de la información.
 - c) Revelar patrones de influencia dentro de una comunidad.
 - d) Ayudar a comprender la conectividad y la accesibilidad entre las entidades.

8. El algoritmo de PageRank puede describirse como un procedimiento utilizado para:
 - a) Determinar la velocidad de carga de una página web en un navegador.
 - b) Calcular la relevancia de una página web en función de la cantidad y calidad de los enlaces que apuntan hacia ella.
 - c) Clasificar las páginas web en función de su edad y autoridad.
 - d) Prevenir el spam y el contenido no deseado en las páginas web.
9. El uso de Hadoop y MapReduce en el contexto de la RI tiene como objetivo:
 - a) Facilitar el análisis en tiempo real de datos de redes sociales.
 - b) Automatizar el mantenimiento de sistemas de bases de datos.
 - c) Mejorar la eficiencia energética en centros de datos.
 - d) Procesar y analizar grandes conjuntos de datos para la RI.
10. Se puede afirmar que:
 - a) La cantidad de conexiones de un nodo siempre indica su influencia en la red.
 - b) Todas las relaciones en una red tienen la misma importancia para el análisis.
 - c) No es posible identificar subgrupos dentro de una red utilizando análisis de redes.
 - d) El tamaño de una red es siempre indicativo de su efectividad en la transmisión de información.
11. La política de ordenación de URLs en los Web Crawlers tiene como aspecto fundamental:
 - a) Limitar el acceso de los crawlers a ciertas secciones de un sitio web, evitando el rastreo de URLs consideradas menos importantes o sensibles.
 - b) Establecer la prioridad de rastreo de las URLs, determinando el orden en que los crawlers visitan y procesan cada página web.
 - c) Definir la estructura de la URL de destino, asegurando que estén ordenadas alfabéticamente para facilitar la navegación y la indexación.
 - d) Determinar la forma en que los crawlers asignan un valor de relevancia a cada URL para clasificarlas en el índice de búsqueda.
12. Con respecto a WordNet y su contribución a las ontologías y la representación del conocimiento en los sistemas de RI, puede afirmarse que:
 - a) En WordNet, una palabra está asociada a un conjunto de sinónimos (synsets), siendo estas palabras intercambiables en un contexto.
 - b) WordNet diferencia claramente entre los significados de palabras según su uso en diferentes contextos, lo que permite aplicaciones avanzadas en desambiguación semántica más allá de los sistemas de recomendación y RI.
 - c) La integración de WordNet en sistemas de RI permite la expansión de consultas y la mejora de la precisión de los resultados al entender mejor el significado de los términos de búsqueda a través de su contexto semántico y las relaciones entre palabras.
 - d) Una limitación de WordNet en la representación del conocimiento es su enfoque en el idioma inglés, lo que plantea desafíos en la aplicación global y la interoperabilidad con sistemas de información multilingües.
13. La centralidad de intermediación de un nodo indica:
 - a) La frecuencia con la que un nodo actúa como puente en el camino más corto entre otros dos nodos.
 - b) El número total de conexiones entrantes y salientes.
 - c) La resistencia del nodo a fallos.
 - d) La cantidad de nodos vecinos directos.
14. Al realizar la optimización de contenido para SEO debe considerarse:
 - a) Utilizar etiquetas de título y meta descripciones únicas y relevantes para cada página.
 - b) Seleccionar las palabras clave al azar.
 - c) Crear contenido valioso y original que satisfaga las necesidades de los usuarios.
 - d) Incluir palabras clave de manera excesiva en el contenido para mejorar el posicionamiento.
15. Para transformar el contenido no estructurado de las páginas web en datos estructurados, el Web Scraping:
 - a) Convierte de forma automática imágenes a texto.
 - b) Interpreta el código JavaScript en tiempo real.
 - c) Extrae información basada en patrones de HTML/CSS.
 - d) Analiza los protocolos de red.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. El componente responsable de la gestión de recursos y planificación de tareas en Hadoop es:
 - a) MapReduce.
 - b) HDFS.
 - c) Hadoop Common.
 - d) YARN.
2. La premisa básica del algoritmo de PageRank para clasificar páginas web en los resultados de búsqueda es:
 - a) Los enlaces entrantes a una página web desde otras páginas contribuyen a su importancia y clasificación.
 - b) La antigüedad de la página web es el principal factor para determinar su clasificación.
 - c) La cantidad de visitas que recibe una página web determina su clasificación.
 - d) El contenido y la relevancia de las palabras clave en la página web determinan su posición.
3. Sobre WordNet y su aplicación en el procesamiento del lenguaje natural, se puede afirmar que:
 - a) Aunque WordNet es una herramienta valiosa en el NLP, su estructura no incluye información sobre la frecuencia de uso de las palabras en el lenguaje natural.
 - b) El diseño de WordNet facilita su integración en aplicaciones multilingües de NLP, aunque su desarrollo original se centró en el inglés.
 - c) Los synsets facilitan la identificación de relaciones semánticas entre palabras, como la hiperonimia y la meronimia, enriqueciendo tareas de NLP.
 - d) WordNet proporciona una base para la desambiguación semántica al agrupar palabras con significados similares en synsets.
4. Una buena práctica de SEO para mejorar el posicionamiento de un sitio web en los motores de búsqueda es:
 - a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Copiar contenido directamente de otros sitios web populares para aumentar la cantidad de páginas indexadas.
 - c) Llenar el contenido con palabras clave irrelevantes para aumentar la densidad de palabras clave.
 - d) Obtener enlaces de otros sitios web relevantes y de calidad que apunten al sitio.
5. Con respecto a WordNet y su contribución a las ontologías y la representación del conocimiento en los sistemas de RI, puede afirmarse que:
 - a) En WordNet, una palabra está asociada a un conjunto de sinónimos (synsets), siendo estas palabras intercambiables en un contexto.
 - b) WordNet diferencia claramente entre los significados de palabras según su uso en diferentes contextos, lo que permite aplicaciones avanzadas en desambiguación semántica más allá de los sistemas de recomendación y RI.
 - c) La integración de WordNet en sistemas de RI permite la expansión de consultas y la mejora de la precisión de los resultados al entender mejor el significado de los términos de búsqueda a través de su contexto semántico y las relaciones entre palabras.
 - d) Una limitación de WordNet en la representación del conocimiento es su enfoque en el idioma inglés, lo que plantea desafíos en la aplicación global y la interoperabilidad con sistemas de información multilingües.
6. En una red de transporte donde cada nodo es una parada de autobús y las aristas representan si existe un carro que pasa por ambos sitios, ¿qué puede mejorar el sistema de transporte?
 - a) Utilizar la centralidad de intermediación para identificar las paradas de transferencia clave y establecer nuevas conexiones entre diferentes líneas de transporte público.
 - b) Utilizar el grafo inducido de los nodos con mayor valor en la centralidad de grado para aplicar la centralidad de intermediación con el propósito de reforzar las paradas con mayor tráfico.
 - c) No implementar sistemas de información en tiempo real para los usuarios, ya que pueden aumentar la carga de los trabajadores.

- d) No considerar la centralidad de intermediación de las estaciones de transporte público al planificar rutas y horarios, ya que no tiene impacto en las conexiones entre las paradas.
7. En el contexto de la RI en redes sociales se puede afirmar que:
- a) El análisis de redes sociales se centra exclusivamente en la cantidad de seguidores que tiene un usuario en particular para determinar su influencia en la red.
 - b) El análisis de redes sociales no es útil para comprender la difusión de información en una red social específica, para ello se utiliza la medida de centralidad de vector propio.
 - c) El análisis de centralidad de intermediación se utiliza para identificar usuarios que son importantes en una red social debido a su posición como “puentes” entre diferentes grupos de usuarios.
 - d) El análisis de sentimientos se utiliza para determinar la popularidad de una publicación en redes sociales sin tener en cuenta la opinión de los usuarios.
8. ¿Qué estrategia utilizan los Web Crawlers para asegurar un rastreo eficiente y respetuoso de los recursos de los sitios web?
- a) Extraer únicamente contenido multimedia para reducir la carga en los servidores web.
 - b) Visitar y rastrear todos los enlaces de una página web simultáneamente.
 - c) Ignorar completamente el archivo `robots.txt` de los sitios web.
 - d) Seguir las directrices del archivo `robots.txt` y aplicar un retraso entre las solicitudes.
9. Analizar una red permite:
- a) Evaluar la calidad del contenido de un sitio web.
 - b) Obtener predicciones exactas de eventos futuros en mercados financieros.
 - c) Detectar posibles tendencias antes de que se conviertan en tendencia.
 - d) Encontrar nodos “sensibles” o críticos para la red.
10. La Web 3 se conoce como:
- a) Web de solo lectura.
 - b) Internet de las cosas.
 - c) Web de escritura-lectura.
 - d) Web semántica.
11. Considerando las dimensiones y desafíos inherentes a Big Data puede afirmarse, tomando en cuenta las características clave y las implicaciones para su procesamiento y análisis, que:
- a) Big Data se caracteriza principalmente por su pequeño volumen y uniformidad, permitiendo un procesamiento eficiente con mínimas adaptaciones de las herramientas de análisis de datos tradicionales.
 - b) Big Data no desafía la capacidad de las herramientas tradicionales de procesamiento de datos para capturar, almacenar, gestionar y analizar efectivamente la información, dada la evolución constante de las capacidades computacionales y algoritmos de optimización.
 - c) Aunque Big Data puede incluir datos estructurados, su naturaleza se expande al incorporar grandes cantidades de datos no estructurados y semiestructurados, lo que exige el uso de tecnologías especializadas en almacenamiento y procesamiento como Hadoop y sistemas de bases de datos NoSQL.
 - d) Además de su complejidad y diversidad, Big Data introduce desafíos significativos en términos de veracidad y variabilidad al requerir métodos avanzados de limpieza y validación de datos para asegurar la integridad del análisis.
12. No se considera como técnica para detectar comunidades en una red:
- a) Analizar la mutualidad de los enlaces.
 - b) Usar el agrupamiento jerárquico.
 - c) Utilizar el algoritmo de K-Means.
 - d) Encontrar cliques de vértices de grado par.
13. El algoritmo Hypertext Induced Topic Selection (HITS) intenta buscar nodos especiales. Estos son conocidos como:
- a) Sensibles.
 - b) Centrales.
 - c) Hubs.
 - d) Autoridades.
14. En el algoritmo de Indexación Basada en Clasificación Bloqueada (BSBI), ¿cuál es el paso final para crear un índice invertido para la colección completa de los datos?
- a) Dividir la colección de datos en bloques de tamaño fijo.
 - b) Eliminar los términos duplicados de los índices.
 - c) Indexar cada bloque de forma independiente.
 - d) Fusionar los índices invertidos de cada bloque.
15. Para contribuir positivamente al posicionamiento orgánico de un sitio web en los motores de búsqueda se puede:

- a) Crear contenido relevante y de alta calidad que satisfaga las necesidades de información de los usuarios.
- b) Obtener enlaces entrantes de otros sitios web de alta autoridad y relevancia temática.
- c) Incluir una densidad alta de palabras clave para asegurar que el sitio web aparezca en tantas búsquedas como sea posible.
- d) Mejorar la velocidad de carga del sitio web y asegurar que sea *responsive* y fácil de usar en dispositivos móviles.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. En el algoritmo de Indexación Basada en Clasificación Bloqueada (BSBI), ¿cuál es el paso final para crear un índice invertido para la colección completa de los datos?
 - a) Dividir la colección de datos en bloques de tamaño fijo.
 - b) Eliminar los términos duplicados de los índices.
 - c) Indexar cada bloque de forma independiente.
 - d) Fusionar los índices invertidos de cada bloque.
2. Considerando las dimensiones y desafíos inherentes a Big Data puede afirmarse, tomando en cuenta las características clave y las implicaciones para su procesamiento y análisis, que:
 - a) Big Data se caracteriza principalmente por su pequeño volumen y uniformidad, permitiendo un procesamiento eficiente con mínimas adaptaciones de las herramientas de análisis de datos tradicionales.
 - b) Big Data no desafía la capacidad de las herramientas tradicionales de procesamiento de datos para capturar, almacenar, gestionar y analizar efectivamente la información, dada la evolución constante de las capacidades computacionales y algoritmos de optimización.
 - c) Aunque Big Data puede incluir datos estructurados, su naturaleza se expande al incorporar grandes cantidades de datos no estructurados y semiestructurados, lo que exige el uso de tecnologías especializadas en almacenamiento y procesamiento como Hadoop y sistemas de bases de datos NoSQL.
 - d) Además de su complejidad y diversidad, Big Data introduce desafíos significativos en términos de veracidad y variabilidad al requerir métodos avanzados de limpieza y validación de datos para asegurar la integridad del análisis.
3. La premisa básica del algoritmo de PageRank para clasificar páginas web en los resultados de búsqueda es:
 - a) Los enlaces entrantes a una página web desde otras páginas contribuyen a su importancia y clasificación.
 - b) La antigüedad de la página web es el principal factor para determinar su clasificación.
 - c) La cantidad de visitas que recibe una página web determina su clasificación.
 - d) El contenido y la relevancia de las palabras clave en la página web determinan su posición.
4. Sobre WordNet y su aplicación en el procesamiento del lenguaje natural, se puede afirmar que:
 - a) Aunque WordNet es una herramienta valiosa en el NLP, su estructura no incluye información sobre la frecuencia de uso de las palabras en el lenguaje natural.
 - b) El diseño de WordNet facilita su integración en aplicaciones multilingües de NLP, aunque su desarrollo original se centró en el inglés.
 - c) Los synsets facilitan la identificación de relaciones semánticas entre palabras, como la hiperonimia y la meronimia, enriqueciendo tareas de NLP.
 - d) WordNet proporciona una base para la desambiguación semántica al agrupar palabras con significados similares en synsets.
5. El concepto de *rank sink* en el algoritmo de PageRank representa páginas web:
 - a) Con una baja calidad de contenido y una cantidad insuficiente de enlaces salientes, lo que las hace menos relevantes en los resultados de búsqueda.
 - b) Con distribuciones uniformes de la puntuación de PageRank.
 - c) Con un alto número de enlaces salientes que no reciben enlaces entrantes, lo que puede afectar negativamente su puntuación de PageRank.
 - d) Que tienen una puntuación de PageRank más alta que otras debido a la manipulación de enlaces entrantes y salientes.
6. Los algoritmos para detectar comunidades en una red intentan:
 - a) Buscar subgrafos tal que no incluyan nodos cuya ausencia desconecte al subgrafo.
 - b) Seleccionar aleatoriamente nodos de alto grado y sus vecinos.
 - c) Buscar conjuntos donde cada nodo de un mismo conjunto tenga características similares al resto de los nodos del conjunto.

- d) Encontrar grupos donde los nodos pertenecientes a los mismos grupos son cercanos bajo cierta métrica y lejanos con respecto a los nodos de otros grupos.
7. ¿Qué es un 'Uniform Resource Locator (URL) Frontier' en el contexto de Web Crawling?
- Una base de datos que almacena URLs únicas identificadas como recursos en la web.
 - Una técnica para filtrar URLs irrelevantes y mejorar la eficiencia del crawling.
 - Un protocolo que define cómo se deben formatear las URLs para el crawling.
 - Una lista prioritaria de URLs que aún no han sido visitadas por el crawler.
8. Una plataforma de comercio electrónico desea mejorar la experiencia del usuario al permitir una navegación más personalizada y contextualizada. Actualmente los usuarios tienen dificultades para encontrar productos relevantes debido a la gran cantidad de opciones disponibles. La empresa está interesada en implementar características de la Web 2.5 y la Web Semántica para abordar este problema. ¿Qué características podrían ayudar para ofrecer una navegación más personalizada y contextualizada?
- La implementación de ontologías y metadatos para enriquecer la descripción de productos y mejorar la precisión de las recomendaciones.
 - La optimización de la velocidad de carga del sitio web para mejorar la experiencia del usuario y reducir el abandono del carrito de compra.
 - La incorporación de los productos en tendencia en el mercado.
 - La integración de redes sociales para permitir la recomendación de productos basada en las preferencias de amigos y contactos.
9. Al realizar la optimización de contenido para SEO debe considerarse:
- Utilizar etiquetas de título y meta descripciones únicas y relevantes para cada página.
 - Seleccionar las palabras clave al azar.
 - Crear contenido valioso y original que satisfaga las necesidades de los usuarios.
 - Incluir palabras clave de manera excesiva en el contenido para mejorar el posicionamiento.
10. Sobre el algoritmo de PageRank, visto en clase, se puede afirmar que:
- Evalúa la importancia de un sitio web en función de la calidad y cantidad de enlaces entrantes que recibe de otros sitios web.
 - Asigna una puntuación baja a los sitios web que contienen muchos enlaces salientes, ya que indica una falta de relevancia.
 - Asigna una puntuación alta a los sitios web que tienen un gran número de enlaces entrantes sin tener en cuenta la calidad de esos enlaces.
 - Solo tiene en cuenta el contenido en un sitio web para determinar su relevancia en los resultados de búsqueda.
11. En un sistema donde el conocimiento está orientado a objetos se puede asegurar que:
- Los marcos y las bandas son las estructuras utilizadas para la representación del modelo.
 - Las funciones de agregación dificultan poder establecer relaciones entre los objetos.
 - Se enfatiza la atención a la información de la cual se extrajo el conocimiento.
 - No es posible definir especializaciones de los objetos de la vida real dentro del sistema.
12. En una empresa comercial, se tiene una red donde los nodos corresponden a los empleados y las aristas representan la frecuencia con la que colaboran en las ventas. Se busca mejorar la colaboración entre los empleados para aumentar las ventas totales, por lo que la directiva debe:
- No tener en cuenta la centralidad de cercanía de los empleados al diseñar estrategias de colaboración, ya que esto podría introducir complicaciones adicionales en el proceso de trabajo.
 - Utilizar la centralidad de cercanía para identificar los empleados menos cercanos a otros en la red y asignarles tareas individuales para evitar posibles conflictos y desacuerdos en el proceso de colaboración.
 - No considerar la centralidad de cercanía de los empleados al diseñar estrategias de colaboración, ya que todos los empleados tienen el mismo acceso a la información y los recursos.
 - Utilizar la centralidad de cercanía para identificar los empleados más cercanos a otros en la red y promover la colaboración entre ellos, facilitando así la comunicación y el intercambio de conocimientos para mejorar las ventas.
13. En un sistema de control de tráfico urbano basado en reglas, ¿cuál de las siguientes reglas sería más efectiva para manejar situaciones de congestión en una intersección?
- Si hay pocos vehículos en la intersección, reducir el tiempo de los semáforos en verde.
 - Si hay muchos vehículos en la intersección, aumentar el tiempo de los semáforos en verde.
 - Si hay un vehículo de emergencia en la intersección, detener todos los demás vehículos.

- d) Si hay muchos peatones cruzando la intersección, reducir el tiempo de los semáforos en rojo.
14. Dentro del análisis de redes, la centralidad de grado mide la importancia de un nodo basándose en:
- a) La cantidad de vecinos del nodo.
 - b) El número de aristas que posee el nodo.
 - c) La cantidad de veces que aparece el nodo en el camino mínimo entre cualquier par de nodos.
 - d) El grado del nodo.
15. En el contexto del procesamiento de grandes conjuntos de datos, se puede asegurar que:
- a) MapReduce no es adecuado para el procesamiento de datos no estructurados.
 - b) MapReduce solo puede manejar pequeñas cantidades de datos y no escala bien a grandes conjuntos de datos.
 - c) MapReduce es un enfoque para procesar datos de forma secuencial en un solo servidor para evitar problemas de concurrencia.
 - d) MapReduce divide una tarea en múltiples pasos de map y reduce que se ejecutan de forma secuencial en diferentes servidores para mejorar el rendimiento y la escalabilidad.



Nombre: _____ Grupo: _____

Seleccione, o no, la(s) respuesta(s) correcta(s) según sea el caso.

1. Se puede afirmar que:
 - a) Uno de los desafíos en el procesamiento de Big Data es la capacidad de gestionar y analizar datos provenientes de diversas fuentes y en diferentes formatos de manera eficiente.
 - b) MapReduce es un modelo de procesamiento distribuido utilizado para trabajar con grandes volúmenes de datos.
 - c) Los SRI pueden beneficiarse de MapReduce para mejorar la RI relevante.
 - d) El término “Big Data” se refiere exclusivamente al volumen de datos que una organización maneja, sin tener en cuenta la velocidad, la variedad y la veracidad de los datos.
2. No se considera como técnica para detectar comunidades en una red:
 - a) Analizar la mutualidad de los enlaces.
 - b) Usar el agrupamiento jerárquico.
 - c) Utilizar el algoritmo de K-Means.
 - d) Encontrar cliques de vértices de grado par.
3. Se tiene un grafo donde cada nodo es un personaje de cierto libro de cuentos y la existencia de las aristas está definida si dos personajes aparecen en el mismo cuento. Se puede asegurar que:
 - a) La centralidad de grado es útil para identificar los personajes con más conexiones dentro de la red, lo que puede indicar su importancia en el libro.
 - b) La centralidad de cercanía indica el grado de conexión de cada personaje con el resto de los personajes de los cuentos del libro.
 - c) La centralidad de intermediación identifica a los personajes que actúan como conectores entre personajes de cuentos distintos.
 - d) La centralidad de cercanía ofrece una relación entre la cantidad de vecinos de un nodo con respecto a la longitud máxima de un camino dentro del grafo partiendo del nodo en cuestión.
4. ¿Qué estrategia utilizan los Web Crawlers para asegurar un rastreo eficiente y respetuoso de los recursos de los sitios web?
 - a) Extraer únicamente contenido multimedia para reducir la carga en los servidores web.
 - b) Visitar y rastrear todos los enlaces de una página web simultáneamente.
 - c) Ignorar completamente el archivo `robots.txt` de los sitios web.
 - d) Seguir las directrices del archivo `robots.txt` y aplicar un retraso entre las solicitudes.
5. La integración de Hadoop y MapReduce en la RI trae como ventaja:
 - a) La posibilidad de la extracción de información relevante y la generación de resultados significativos de grandes conjuntos de datos.
 - b) La reducción de los costos operativos a cero.
 - c) La eliminación de la necesidad de sistemas de bases de datos.
 - d) La garantía de la privacidad absoluta de los datos procesados.
6. En el modelo de representación del conocimiento basado en herencia se puede asegurar que:
 - a) La cantidad de padres que puede tener un nodo no es mayor que 4.
 - b) Solo se usa en entornos referentes a la biología.
 - c) El razonamiento deducido está respaldado por al menos un camino dentro del grafo.
 - d) Las conclusiones pueden ser canceladas si el grafo es ambiguo.
7. ¿Qué es un 'Uniform Resource Locator (URL) Frontier' en el contexto de Web Crawling?
 - a) Una base de datos que almacena URLs únicas identificadas como recursos en la web.
 - b) Una técnica para filtrar URLs irrelevantes y mejorar la eficiencia del crawling.
 - c) Un protocolo que define cómo se deben formatear las URLs para el crawling.
 - d) Una lista prioritaria de URLs que aún no han sido visitadas por el crawler.
8. Una buena práctica de SEO para mejorar el posicionamiento de un sitio web en los motores de búsqueda es:

- a) Utilizar técnicas de encubrimiento para mostrar contenido diferente a lo indexado por los motores de búsqueda.
 - b) Copiar contenido directamente de otros sitios web populares para aumentar la cantidad de páginas indexadas.
 - c) Llenar el contenido con palabras clave irrelevantes para aumentar la densidad de palabras clave.
 - d) Obtener enlaces de otros sitios web relevantes y de calidad que apunten al sitio.
9. Analizar una red permite:
- a) Evaluar la calidad del contenido de un sitio web.
 - b) Obtener predicciones exactas de eventos futuros en mercados financieros.
 - c) Detectar posibles tendencias antes de que se conviertan en tendencia.
 - d) Encontrar nodos “sensibles” o críticos para la red.
10. En un sistema donde el conocimiento está orientado a objetos se puede asegurar que:
- a) Los marcos y las bandas son las estructuras utilizadas para la representación del modelo.
 - b) Las funciones de agregación dificultan poder establecer relaciones entre los objetos.
 - c) Se enfatiza la atención a la información de la cual se extrajo el conocimiento.
 - d) No es posible definir especializaciones de los objetos de la vida real dentro del sistema.
11. En la RI en el contexto de Big Data, se puede asegurar que:
- a) El almacenamiento distribuido es una técnica obsoleta.
 - b) El almacenamiento distribuido centraliza todos los datos en un único servidor para facilitar su gestión y mantenimiento.
 - c) El almacenamiento distribuido reparte los datos en múltiples servidores para mejorar la disponibilidad y la redundancia del sistema.
 - d) El almacenamiento distribuido ofrece ventajas significativas en términos de escalabilidad y rendimiento en comparación con el almacenamiento centralizado.
12. ¿Cuál de las siguientes opciones describe mejor la diferencia clave entre Web Crawling y Web Scraping?
- a) Web Scraping implica el análisis de la estructura y el contenido de las páginas web para extraer datos, mientras que Web Crawling se refiere a la descarga y almacenamiento de páginas web completas.
 - b) Web Scraping es más eficaz para rastrear e indexar contenido web para motores de búsqueda, mientras que Web Crawling se utiliza principalmente para la extracción de datos en proyectos de investigación.
 - c) Web Crawling se centra en la exploración y recopilación de enlaces de múltiples sitios web, mientras que Web Scraping se enfoca en la extracción específica de datos de páginas web individuales.
 - d) Web Crawling se realiza utilizando herramientas de automatización como Selenium WebDriver, mientras que Web Scraping se lleva a cabo mediante el análisis de HTML y CSS.
13. En el contexto del procesamiento de grandes conjuntos de datos, se puede asegurar que:
- a) MapReduce no es adecuado para el procesamiento de datos no estructurados.
 - b) MapReduce solo puede manejar pequeñas cantidades de datos y no escala bien a grandes conjuntos de datos.
 - c) MapReduce es un enfoque para procesar datos de forma secuencial en un solo servidor para evitar problemas de concurrencia.
 - d) MapReduce divide una tarea en múltiples pasos de map y reduce que se ejecutan de forma secuencial en diferentes servidores para mejorar el rendimiento y la escalabilidad.
14. El algoritmo de PageRank puede describirse como un procedimiento utilizado para:
- a) Determinar la velocidad de carga de una página web en un navegador.
 - b) Calcular la relevancia de una página web en función de la cantidad y calidad de los enlaces que apuntan hacia ella.
 - c) Clasificar las páginas web en función de su edad y autoridad.
 - d) Prevenir el spam y el contenido no deseado en las páginas web.
15. Al “relajar” el concepto de clique en la detección de comunidades se intenta solucionar:
- a) La uniformidad de los nodos en términos de grado.
 - b) La necesidad de datos externos para analizar la red.
 - c) La imposibilidad de conectar nodos distantes.
 - d) El solapamiento y la complejidad computacional.