Actividad 5: Caso de estudio NLP en e-commerce

1. Investigar en qué consiste faceted search y semantic search.
   1. Faceted search: es una variante especializada de búsqueda que permite a un usuario buscar por una gran cantidad de filtros. Por ejemplo, si un usuario quiere comprar un televisor, la plataforma ofrece filtros personalizados del producto que está buscando en el lado izquierdo de la página web
   2. Semantic search: Los usuarios utilizan lenguaje natural para buscar artículos en una web de e-commerce. El lenguaje natural es difícil de entender para los motores de búsqueda y no puede diferenciar entre nombres y descripciones de productos. La búsqueda semántica puede identificar errores tipográficos, términos de búsqueda más largos e incluso reconocer sinónimos. La búsqueda semántica también puede analizar el historial de búsqueda y predecir los términos que escribe el usuario. El autocompletado ahorra tiempo a los clientes y les ayuda a encontrar lo que buscan más rápido. El procesamiento del lenguaje natural permite la búsqueda inteligente para comprender y también consultar contenido digital de diversas fuentes de datos ayudando a la búsqueda inteligente a desglosar términos lingüísticos, sinónimos y cualquier relación en el lenguaje cotidiano
2. Indicar cómo se aplican faceted search y semantic search en e-commerce para buscar ítems en un marketplace y como una complementa a la otra.
   1. **Faceted Search** y **Semantic Search** se complementan de manera que el primero permite a los usuarios refinar manualmente los resultados en función de atributos específicos, mientras que el segundo ofrece resultados más relevantes y contextualizados basados en la interpretación de la intención de búsqueda.
3. Buscar otros dominios donde se aplique faceted search y semantic search. Explicar cómo se aplican en el dominio elegido.
   1. Bibliotecas digitales y repositorios académicos: Faceted Search permite a los usuarios filtrar libros, artículos y otros recursos por categorías como autor, año de publicación, tema, tipo de documento, revista. Semantic Search se utiliza para interpretar el significado del contenido, facilitando la búsqueda de documentos académicos relevantes basados en la intención de la consulta.
   2. Salud y medicina: FS puede buscar información médica, tratamientos, o ensayos clínicos utilizando filtros como tipo de enfermedad, tipo de tratamiento, edad del paciente, resultados de tratamientos. SS permite realizar consultas complejas interpretando términos médicos y sinónimos.
   3. Sector financiero: pueden utilizar Faceted Search para buscar productos financieros como cuentas bancarias, préstamos, tarjetas de crédito, y seguros, filtrando por tasa de interés, tipo de producto, período, beneficios adicionales. El Semantic Search en el sector financiero permite a los usuarios realizar consultas complejas como "cuentas de ahorro con los mejores intereses para jóvenes" o "préstamos para la compra de vivienda", y recibir resultados relevantes que interpreten la intención y las necesidades específicas.
4. Investigar diferentes técnicas modernas para extraer atributos para el armado del catálogo de un e-commerce, armar una síntesis de cada una y una tabla comparativa de las técnicas encontradas.
   1. Direct attribute extraction: se modelan como un problema sequence-to-sequence supervisado (Token classification). Como input se recibe una secuencia de texto y como output devuelve una secuencia de texto etiquetada. Esto se realiza entrenando un modelo como un named entity recognizer. Esto se puede trabajar con expresiones regulares para poder encontrar especificaciones e inclusive usar categorías más específicas como memoria o almacenamiento según corresponda. Existen varios tipos de features.
      1. Characteristic features: Estas features a nivel de token, el tipo de case del token, lengths y su composición de caracteres
      2. Locational features: Estas features capturan información de la posición del token en la secuencia, como la cantidad de tokens antes del dado o el radio de la posición del token y el total de la secuencia.
      3. Contextual features: Este tipo de features encodean la información sobre los vecinos del token dado, como la identidad del token precedente y el siguiente, POS tag del token, conjunción de los mismos
   2. Indirect attribute extraction: Los atributos indirectos son atributos que no se mencionan en la descripción. Estos atributos son inferidos de otros atributos. Para clasificación de atributos indirectos se usa text classification para inferir clases de alto nivel por ejemplo para inferir géneros. Estas técnicas se pueden complementar con técnicas multimodales para poder analizar texto de descripciones, imágenes y reviews de productos
5. Investigar técnicas modernas para categorizar productos y crear taxonomías para un e-commerce, armar una síntesis de cada una y una tabla comparativa de las técnicas encontradas.
   1. Modelos Supervisados: Se entrenan modelos de machine learning (como Random Forests, SVM, o Redes Neuronales) para clasificar productos en categorías predefinidas utilizando características del producto, como el título, descripción, precio, y etiquetas.
   2. Procesamiento del Lenguaje Natural (NLP): Se utilizan técnicas de NLP para analizar el texto relacionado con los productos (títulos, descripciones, reseñas) y extraer características que ayudan a categorizar los productos. Modelos como BERT o GPT pueden ser empleados para comprender el contexto y semántica del producto.
   3. Ontologías: Se utilizan ontologías para definir relaciones jerárquicas y semánticas entre diferentes categorías de productos. Esto permite crear una estructura de taxonomía que refleje el conocimiento experto en el dominio.
   4. Grafos de Conocimiento: Implementar grafos de conocimiento (knowledge graphs) permite modelar productos, categorías y sus relaciones de manera compleja y dinámica.
   5. K-Means Clustering: Agrupa productos en categorías basadas en similitudes en sus características
6. Investigar diferentes técnicas y enfoques modernos para recomendar productos en un e-commerce, armar una síntesis de cada una y una tabla comparativa de las técnicas encontradas.
   1. Collaborative Filtering Basado en Usuarios: Recomienda productos basándose en la similitud de comportamiento entre usuarios.
   2. Redes Neuronales Convolucionales (CNNs) y Recurrentes (RNNs): Se utilizan para capturar patrones complejos en los datos, como secuencias de acciones de los usuarios (por ejemplo, productos vistos y comprados) y para entender relaciones profundas entre productos.
   3. Autoencoders: Se emplean para reducir dimensionalidad y capturar características latentes que pueden ser útiles para la recomendación.
   4. Análisis de Reseñas de Productos: Se analizan las opiniones de los usuarios sobre productos para identificar aquellos con sentimientos positivos y recomendarlos.
   5. Chatbots y Asistentes Virtuales: Utilizando modelos de procesamiento del lenguaje natural (NLP), como GPT o BERT, los chatbots y asistentes virtuales pueden interactuar con los usuarios en lenguaje natural para ofrecer recomendaciones personalizadas. Estos sistemas entienden las consultas del usuario y sugieren productos relevantes en tiempo real.