

UNIDAD DIDÁCTICA

COMPETENCIAS DIGITALES

Tema

**UTILIZACIÓN DE LA HERRAMIENTA DE
BÚSQUEDA AVANZADA PARA ANÁLISIS
TEXTUAL**

UTILIZACIÓN DE LA HERRAMIENTA DE BÚSQUEDA AVANZADA PARA ANÁLISIS TEXTUAL

¿Qué es el análisis de textos?

El análisis de texto es el proceso en el cual se utilizan los sistemas de computación para leer y comprender texto escrito por seres humanos y, así, obtener información empresarial. El *software* de análisis de datos puede, de forma independiente, clasificar, ordenar y extraer información de distintos textos para identificar patrones, relaciones, opiniones y otra información que se pueda procesar. Puede utilizar el análisis de texto para procesar de manera eficiente y precisa varios orígenes basados en texto, como emails, documentos, contenido de redes sociales y reseñas de productos, tal como lo haría una persona.

¿Por qué es importante el análisis de textos?

Las empresas utilizan el análisis de textos para extraer información procesable de varios orígenes de datos sin estructura. Dependen de la retroalimentación de orígenes como *emails*, redes sociales y respuestas de encuestas a clientes para ayudar en la toma de decisiones. Sin embargo, el inmenso volumen de textos de estos orígenes resulta abrumador sin un *software* de analítica de textos.

Gracias al análisis de textos, puede obtener información precisa de los orígenes más rápidamente. El proceso está completamente automatizado, es constante y muestra datos sobre los que puede actuar. Por ejemplo, el uso del *software* de análisis de textos

permite detectar de inmediato sentimientos negativos en las publicaciones de las redes sociales para que pueda resolver el problema.

Análisis de sentimientos

El análisis de sentimientos o la minería de opiniones utiliza métodos de análisis de textos para comprender la opinión transmitida en un texto. Puede utilizar el análisis de sentimientos de reseñas, blogs, foros y otros medios digitales para determinar si sus clientes están satisfechos con sus compras. El análisis de sentimientos ayuda a detectar tendencias nuevas, realizar un seguimiento de los cambios de sentimientos y abordar problemas de relaciones públicas. Mediante el análisis de sentimientos y la identificación de palabras clave específicas, puede realizar un seguimiento de los cambios en la opinión del cliente e identificar la causa raíz del problema.

Administración de registros

El análisis de textos conduce a la administración, la categorización y la búsqueda eficientes de documentos. Esto incluye la automatización de la administración de registros de pacientes, la supervisión de menciones de la marca y la detección de fraudes de seguros. Por ejemplo, LexisNexis Legal & Professional utiliza la extracción de textos para identificar registros específicos entre 200 millones de documentos.

Personalización de la experiencia del cliente

Puede utilizar el *software* de análisis de textos para procesar *emails*, reseñas, chats y otra correspondencia basada en texto. Gracias a la información sobre las preferencias, los hábitos de compra y la percepción general de la marca de los clientes, usted puede adaptar experiencias personalizadas para diferentes segmentos de clientes.

¿Cómo funciona el análisis de textos?

El núcleo del análisis de textos es el entrenamiento del software de la computadora para que asocie palabras con significados específicos y comprenda el contexto semántico de datos no estructurados. Esto es similar a cómo los humanos aprenden un idioma nuevo y asocian palabras con objetos, acciones y emociones.

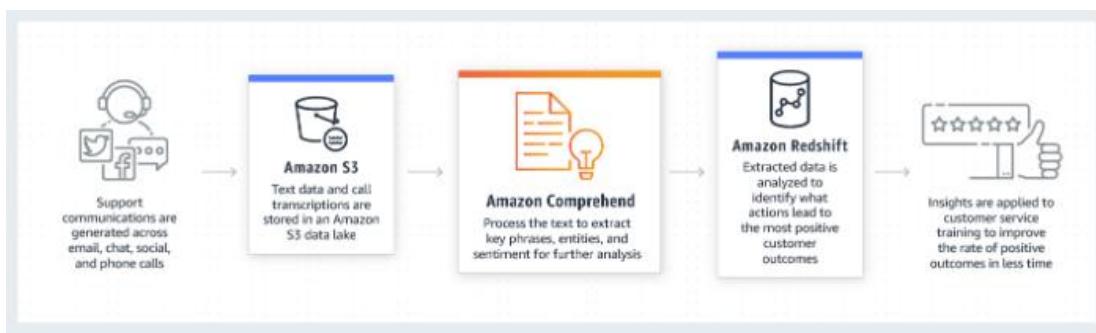
El software de análisis de textos funciona según los principios del aprendizaje profundo y el procesamiento de lenguaje natural.

Aprendizaje profundo

La inteligencia artificial es el campo de la ciencia de datos que enseña a las computadoras a pensar como los humanos. El machine learning es una técnica dentro de la inteligencia artificial que utiliza métodos específicos para enseñar o entrenar a las computadoras. El aprendizaje profundo es un método de machine learning altamente especializado que utiliza redes neuronales o estructuras de software que imitan al cerebro humano. La tecnología de aprendizaje profundo impulsa el software de análisis de textos para que estas redes puedan leer textos de manera similar al cerebro humano.

Procesamiento del lenguaje natural

El procesamiento de lenguaje natural (NLP) es una rama de la inteligencia artificial que proporciona a las computadoras la capacidad de derivar de forma automática el significado del texto natural creado por humanos. Utiliza modelos lingüísticos y estadísticas para entrenar la tecnología de aprendizaje profundo y así, procesar y analizar los datos del texto, incluidas las imágenes del texto escrito a mano. Los métodos de NLP, como el reconocimiento óptico de caracteres (OCR), convierten imágenes de texto en documentos de texto, ya que encuentran y comprenden las palabras en las imágenes.



¿Cuáles son los tipos de técnicas de análisis de textos?

El software de análisis de textos utiliza estas técnicas comunes.

Clasificación de textos

En la clasificación de textos, el *software* de análisis de textos aprende a asociar ciertas palabras clave con temas específicos, intenciones o sentimientos de los usuarios. Lo logra mediante los siguientes métodos:

- La clasificación basada en reglas asigna etiquetas al texto en función de reglas predefinidas para componentes semánticos o patrones sintácticos.
- Los sistemas basados en *machine learning* entran el *software* de análisis de textos con ejemplos y aumentan la precisión del etiquetado del texto. Utilizan modelos lingüísticos como Naive Bayes, máquinas de vectores soporte y aprendizaje profundo para procesar datos estructurados, categorizar palabras y desarrollar una comprensión semántica entre ellos.

Por ejemplo, una reseña favorable habitualmente contiene palabras como *bueno*, *rápido* y *excelente*. Sin embargo, las reseñas negativas pueden contener palabras como *infeliz*, *lento* y *malo*. Los científicos de datos entran el *software* de análisis de textos para buscar términos específicos y categorizar las reseñas como positivas o negativas. De esta manera, el equipo de atención al cliente puede supervisar fácilmente los sentimientos de los clientes a partir de las reseñas.

Extracción de texto

La extracción de texto escanea el texto y extrae información clave. Puede identificar palabras clave, atributos de los productos, nombres de marcas, nombres de lugares y más en un texto. El *software* de extracción aplica los siguientes métodos:

- Expresión regular (REGEX): esta es una matriz formateada de símbolos que sirve como condición previa de lo que se debe extraer.
- Campos aleatorios condicionales (CRF): este es un método de *machine learning* que extrae textos mediante la evaluación de patrones o frases específicos. Es más refinado y flexible que REGEX.

Por ejemplo, la extracción de texto se puede utilizar para supervisar las menciones de la marca en las redes sociales. Es imposible hacer un seguimiento manual de cada aparición de la marca en las redes sociales. La extracción de texto alertará las menciones de su marca en tiempo real.

Modelado de temas

Los métodos de modelado de temas identifican y agrupan palabras clave relacionadas que aparecen en un tema en un texto no estructurado. Estos métodos pueden leer varios documentos de texto y ordenarlos por temas en función de la frecuencia de varias palabras en el documento. Los métodos de modelado de temas proporcionan contexto para un análisis más profundo de los documentos.

Por ejemplo, puede utilizar métodos de modelado de temas para leer su archivo de documento escaneado y clasificar documentos en facturas, documentos legales y acuerdos con clientes. Luego, puede ejecutar diferentes métodos de análisis en las facturas para obtener información financiera o en los acuerdos con clientes para obtener información sobre el cliente.

Redacción de información de identificación personal (PII)

La redacción de PII detecta y elimina de forma automática la información de identificación personal (PII) como nombres, direcciones o números de cuenta de un documento. La redacción de PII ayuda a proteger la privacidad y cumplir con las leyes y las regulaciones locales.

Por ejemplo, puede analizar tickets de soporte y artículos de conocimiento para detectar y redactar PII antes de indexar los documentos a la solución de búsqueda. Después de eso, las soluciones de búsqueda no tienen PII en los documentos.

¿Cuáles son las etapas del análisis de textos?

Para implementar el análisis de textos, debe seguir un proceso sistemático que pasa por cuatro etapas.

Etapa 1: Recopilación de datos

En esta etapa, se recopilan datos de texto de orígenes internos o externos.

Datos internos

Los datos internos hacen referencia al contenido de texto interno de la empresa y se puede acceder a estos con facilidad. Por ejemplo, correos electrónicos, chats, facturas y encuestas de empleados.

Datos externos

Puede encontrar datos externos en orígenes como publicaciones de redes sociales, reseñas en línea, artículos nuevos y foros en línea. Es más difícil adquirir datos externos porque están fuera de su control. Es posible que tenga que utilizar herramientas de raspado web o integrar soluciones de terceros para extraer datos externos.

Etapa 2: Preparación de los datos

La preparación de datos es una parte esencial del análisis de textos. Supone estructurar datos de texto sin procesar en un formato aceptable para el análisis. El *software* de análisis de textos automatiza el proceso e involucra los siguientes métodos comunes de procesamiento de lenguaje natural (NLP).

Tokenización

La tokenización es la segregación del texto sin procesar en varias partes que tienen sentido semántico. Por ejemplo, la frase *los análisis de textos beneficia a las empresas* tokeniza las palabras *texto, análisis, beneficia y empresas*.

Etiquetado de partes del discurso

El etiquetado de partes del discurso asigna etiquetas gramaticales al texto tokenizado. Por ejemplo, aplicar este paso a los tokens mencionados anteriormente da como resultado *texto: Sustantivo; análisis: Sustantivo; beneficia: Verbo; empresas: Sustantivo*.

Análisis sintáctico

El análisis sintáctico establece conexiones significativas entre las palabras tokenizadas con la gramática inglesa. Ayuda al *software* de análisis de textos a visualizar la relación entre las palabras.

Lematización

La lematización es un proceso lingüístico que simplifica las palabras en su forma de diccionario o lema. Por ejemplo, *visualización* aparece en el diccionario como *visualizar*.

Eliminación de palabras vacías

Las palabras vacías son palabras que ofrecen poco o ningún contexto semántico a una oración, como *y, o* y *para*. Según el caso de uso, el *software* puede eliminarlas del texto estructurado.

Etapa 3: Análisis de textos

El análisis de textos es la parte fundamental del proceso, en el que el *software* de análisis de textos procesa el texto mediante diferentes métodos.

Clasificación de textos

La clasificación es el proceso de asignación de etiquetas a los datos de texto que se basan en reglas o en sistemas basados en *machine learning*.

Extracción de texto

La extracción supone identificar la presencia de palabras clave específicas en el texto y asociarlas con etiquetas. El *software* utiliza métodos como expresiones regulares y campos aleatorios condicionales (CRF) para lograrlo.

Etapa 4: Visualización

La visualización se trata de convertir los resultados del análisis de textos en un formato fácil de comprender. Encontrará resultados de analítica de textos en gráficos, cuadros y tablas. Los resultados visualizados lo ayudan a identificar patrones y tendencias, y crear los planes de acción correspondientes. Por ejemplo, imagine que está experimentando un aumento en las devoluciones de productos, pero tiene problemas para identificar las causas. Con la visualización, se buscan palabras como *defectos*, *tamaño incorrecto* o *no encaja bien* en los comentarios y se tabulan en un gráfico. Entonces sabrá cuál es el problema principal que tiene más prioridad.

¿Qué es la analítica de textos?

La analítica de textos trata de los datos cuantitativos que se pueden obtener mediante el análisis de patrones en varias muestras del texto. Se presenta en cuadros, tablas o gráficos.

Análisis de textos frente a analítica de textos

La analítica de textos ayuda a determinar si existe una tendencia o un patrón en particular a partir de los resultados del análisis de miles de comentarios. Mientras tanto, puede utilizar el análisis de textos para determinar si los comentarios de un cliente son positivos o negativos.

¿Qué es la minería de textos?

La minería de textos se trata del proceso de obtención de información cualitativa mediante el análisis de textos no estructurados.

Análisis de datos frente a minería de textos

No hay diferencia entre el análisis de textos y la minería de textos. Ambos términos se refieren al mismo proceso de obtención de información valiosa de orígenes como el *email*, las respuestas de encuestas y las publicaciones en las redes sociales.

¿Cómo puede ayudar Amazon Comprehend?

Amazon Comprehend es un servicio de procesamiento de lenguaje natural que utiliza machine learning para descubrir información y conexiones valiosas en los textos. Puede utilizarlo para simplificar los flujos de trabajo de procesamiento de documentos mediante la clasificación y la extracción de información automáticas. Por ejemplo, puede utilizar Amazon Comprehend para realizar las siguientes tareas:

- Realice análisis de sentimientos de los tickets de atención al cliente, las reseñas de productos, las publicaciones en las redes sociales, etc.
- Integre Amazon Comprehend con Amazon Lex para desarrollar un chatbot de conversación inteligente.
- Extraiga términos médicos a partir de documentos e identifique la relación entre estos con Amazon Comprehend Medical.



