

Generación automática de contenido a través de la inteligencia artificial

Autores: Alejandra López Ocampo

Facultad de ingeniería en sistemas y computación, UTP, Pereira, Colombia

Correo-e: alejandra.lopez1@utp.edu.co

Resumen— En este texto se tiene como objetivo el explicar y cómo funciona una de las aplicaciones que tiene la rama de la inteligencia artificial que está siendo implementada y que genera bastante controversia como lo es la generación automática de contenidos, además de presentar un modelo modular con el que se está llevando a cabo para la redacción de contenido a tiempo real usando como herramienta el web scraping y la búsqueda por palabras clave.

Palabras clave – Inteligencia artificial, Web scraping, categorías, palabras clave, documento maestro.

Abstract— This text aims to explain and how one of the applications that has the branch of artificial intelligence that is being implemented and that generates a lot of controversy such as automatic content generation, in addition to presenting a modular model with the which is being carried out for writing content in real time using web scraping and keyword searching as a tool.

Key Word — Artificial intelligence, Web scraping, categories, keywords.

I. INTRODUCCION

En primer lugar, es importante destacar que gracias a la inteligencia artificial se han podido automatizar varias tareas que antes no se creía posible. De la misma manera desde hace ya varios años se está intentando automatizar la creación de contenido haciendo que sea posible que mediante software se puedan producir piezas de información coherentes a tal punto de que el lector no sea capaz de vislumbrar la diferencia entre si escrito por una AI o por un escritor.

Asimismo, uno de los principales usos que se le da a esta es la generación de artículos, blogs, ensayos y demás tipos de contenido.

II. CONTENIDO

La generación de contenido automática ya ha sido implementada anteriormente, muchos de los grandes en la industria del periodismo y muchos portales en la web hacen uso de esta tecnología.

Creación automática de contenido en la actualidad:

Heliograph:

Esta es una herramienta del Washington Post la cual realizó su debut durante las olimpiadas de Brasil en el 2016. Esta redactaba artículos cortos con información de la competencia. Actualmente está siendo utilizado para cubrir noticias locales. Su principal enfoque está siendo en que no solo se generen contenidos y se compartan estadísticas si no también que sea en tiempo real.

Automated Insights:

Este sistema es la que ayuda a The associated press a aumentar la producción de notas y artículos. Generalmente son trabajos informativos breves donde redacta un modelo de artículo, el cual puede ser visto y editado

Online Content Toolkit (OCTO):

Esta herramienta está aún en desarrollo es un proyecto de la BBC que está enfocado a explorar la edición de contenido multimedia mediante el análisis inteligente.

Puede definirse como un conjunto de herramientas que a su vez pueden usarse individualmente como, por ejemplo, se tiene algo denominado como “speech recognition” o reconocimiento de voz, el cual tiene como objetivo traducirlos a texto.

En el diagrama siguiente se podrá ver como funciona y como está estructurado este proyecto.

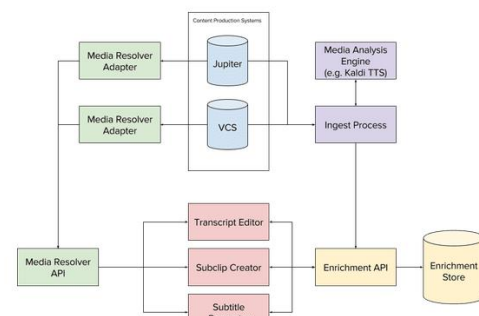


Figura 1. Como funciona el proyecto “Online Content Toolkit (OCTO)”

Herramientas online:

- Article Generator
- Automatic Text Summarizer
- Article Creator

Categorización usando la inteligencia artificial:

Esta es una fase clave para este tipo de sistemas ya que cada artículo, ensayo o contenido generado pertenece a un tema específico que está dentro de alguna categoría, con lo cual para poder redactar el texto deberá realizar una búsqueda de información que está indexada por tema, subtema, hasta llegar al que se requiere específicamente.

Para lo anterior es requerida una fuente de información que contenga los temas categorizados

Web Scraping

El web scraping es usado para extraer lo que se conoce como “scrape” de cualquier portal de internet. Estos procesos simulan la navegación normal de un humano en la web y está altamente ligado a la indexación, ya que mediante esta funcionan la mayoría de motores de búsqueda. Esta se suele implementar bajo lenguajes como PHP, Java y demás.

Esta técnica es muy útil para poder implementar el modelo ya que al igual que es usada por las páginas web dinámicas

Indexación

En el siguiente gráfico se puede ver cómo está indexado el portal de EL ESPECTADOR, el cual se tomará como referencia ya que puede ser un usuario potencial de este tipo de tecnologías, además de ser uno de los medios con mayor reconocimiento en el país.

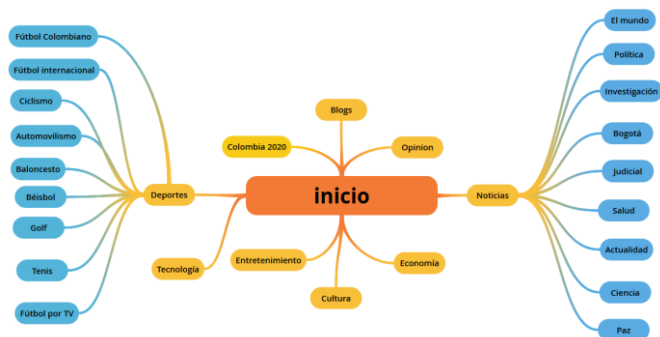


Figura 2. Clasificación de contenido en el portal del espectador

En este apartado podemos visualizar lo importante que es tener la información bien clasificada, no solo para que el

usuario pueda accederla de una manera efectiva, sino también para que el web scrapping sea exitoso.

Librerías de scrapping

Actualmente existen varias librerías que nos permiten

- Pyspider
- MechanicalSoup
- Scrapy
- JSoup

Diccionarios

La construcción de diccionarios es importante debido a que allí es donde se consultan las relaciones con los tópicos que fueron clasificados mediante la indexación, a partir de aquí se podrá hacer una búsqueda usando las palabras clave que le sean proporcionadas para poder comenzar a redactar el documento maestro.

Búsqueda

Para poder crear el documento maestro debe apoyarse en sus diccionarios y las relaciones

Implementación

Pero para la implementación se deberán cargar los diccionarios y además se deberá crear un formato de documento maestro mediante el cual se podrá comenzar a entrenar el modelo hasta llegar a generar un texto coherente basado en las palabras clave.

III. CONCLUSIONES

Es posible concluir que muchas de las herramientas nombradas anteriormente son una gran ayuda para los redactores ya que les evita tener que realizar manualmente trabajos simples que les podrían robar tiempo que puede ser usado para generar otro tipo de contenidos.

Aunque en este campo hay mucho más que solo la generación de textos también está el sistema de recomendaciones inteligentes, el email marketing, chatbots y muchas otras que pueden fortalecer mucho a la industria en general.

REFERENCIAS

- [1] <https://obsbusiness.school/int/blog-investigacion/marketing-y-comunicacion/inteligencia-artificial-en-la-creacion-de-contenidos-mas-que-robots>
- [2] <https://www.postedin.com/blog/los-robots-de-creacion-de-contenido-han-llegado/>

- [3] <https://seoduccion.com/ia-y-creacion-de-contenidos/>
- [4] <https://www.cyberclick.es/numerical-blog/la-inteligencia-artificial-en-el-marketing-digital>
- [5] <https://www.elspectador.com/noticias/>
- [6] <https://www.washingtonpost.com/pr/wp/2016/10/19/thewashington-post-uses-artificial-intelligence-to-cover-nearly-500-races-on-election-day/>
- [7] <https://bbcnewslabs.co.uk/projects/octo/>
- [8] <https://bbcnewslabs.co.uk/static/655bcd0129a3e97f3a478fdf38d11da/b7c40/OCTOdiagram.png>