

Procesamiento y Gestión Automatizada de Archivos Multimedia

Automated Processing and Management of Multimedia Files

Carlos Rojas¹

c.rojas@outlook.com

(1) Facultad de Ingeniería Electrónica e Informática, Universidad Nacional Federico Villarreal, Perú

Resumen

El propósito de esta investigación se centra principalmente en desarrollar un sistema automatizado de procesamiento y gestión de videos que permita a los usuarios tener una plataforma digital donde puedan filtrar y buscar el contenido multimedia de forma eficiente. Se utiliza Inteligencia Artificial y Aprendizaje Automático para entender el contenido de los videos y crear patrones de comportamiento de los usuarios.

Para la solución se usa Servicios Cognitivos de Azure (SCA) para automatizar la extracción de las características de voz a texto, Servicios de Funciones de Azure (SFA) para el diseño elástico y modular de la arquitectura y Machine Learning (ML) para entender el contenido de video y crear patrones de comportamiento de los usuarios.

Palabras clave: *automatización, Azure, computación en la nube, filtrar, procesamiento de videos, servicios cognitivos.*

Abstract

The purpose of this research is mainly to develop an automated system of video processing and management that allows users to have a digital platform where they can filter and search multimedia content efficiently. Artificial Intelligence and Machine Learning are used to understand the content of the videos and create patterns of user's behavior.

Azure Cognitive Services (SCA) is used to automate the extraction of voice-to-text features, Azure Function Services (SFA) for the elastic and modular design of the architecture and the learning machine (ML) for the Video Content and create patterns of user behavior.

Keywords: *automation, Azure, cloud computing, cognitive services, filtering, video processing.*

1. Introducción

Internet fue diseñado alrededor de documentos basados en texto, y como tal, tiene una infraestructura madura para fomentar y permitir la búsqueda y descubrimiento de este en toda la web (Shuler, 2002). Los archivos de video, por otro lado, no se pueden buscar de forma nativa y, por lo general, se requiere complejos sistemas de clasificación que funcionan principalmente con grandes cantidades de metadatos etiquetados manualmente.

Debido a la complejidad y la cantidad de información en los videos y las numerosas tareas requeridas para el proceso de Inteligencia Artificial, el desarrollo y la implementación deben ser altamente modular y utilizar servicios de computación en la nube de nivel empresarial.

Para entender el contenido de los videos y crear patrones de comportamiento de los usuarios se ha utilizado Aprendizaje Automático e Inteligencia Artificial, para automatizar la extracción de las características de voz a texto (Servicios Cognitivos de Azure - SCA), para el diseño elástico y modular de la arquitectura (Servicios de Funciones de Azure -SFA) y para entender el contenido de video y crear patrones de comportamiento de los usuarios (Machine Learning - ML).

Se decidió utilizar Microsoft Azure para automatizar la extracción de las características de voz a texto y luego procesar estos resultados utilizando los Servicios Cognitivos (Microsoft, 2019b). Para el diseño elástico y modular de la arquitectura se decidió utilizar el servicio Functions de Microsoft Azure. El resultado puede ser procesado por un algoritmo de clasificación. Con todo esto logramos definir y construir un proceso automatizado, predecible y elástico.

Por lo tanto, el principal objetivo fue desarrollar un sistema automatizado de procesamiento y gestión de videos que permita tener una plataforma digital donde puedan filtrar y buscar el contenido multimedia de forma eficiente.

2. Material y Método

Se utilizó Inteligencia Artificial y Aprendizaje Automático para entender el contenido de video y crear patrones de comportamiento de los usuarios. También se utilizaron algoritmos especializados para crear grupos de videos y dar recomendaciones. Se ha creado ontologías personalizadas para cada conjunto de datos de video, lo que permite la búsqueda avanzada de gran alcance en el contenido del video.

Se decidió utilizar diversos servicios de Microsoft Azure para importantes tareas como automatizar la extracción de las características de voz a texto y luego procesar estos resultados; así como para el diseño elástico y modular de la arquitectura. El resultado puede ser procesado por un algoritmo de clasificación. Con todo esto logramos definir y construir

un proceso automatizado, predecible y elástico. Las tecnologías utilizadas en este proyecto son: Azure Cognitive Services, Azure Media Services, Azure App Service, Azure Search, Azure Storage y Application Insights; y las herramientas de desarrollo de programación Visual Studio, Visual Studio Code, C# (.NET), Azure CLI, Azure PowerShell, Visual Studio Team Services. En la figura 2 mostramos los principales servicios de Azure usados en la solución, y en la figura 3 mostramos la estructura del código en Visual Studio.

También se usó de Azure Media Services para extraer características básicas para luego procesarlas con Azure Cognitive Services para obtener información útil para clasificar y recomendar videos.

Durante este desarrollo, hemos observado más de cerca y de primera mano las capacidades de los Servicios de Aprendizaje Automático para la indexación de audio y los Servicios Cognitivos más recientes, los Servicios de Colas y el diseño de serveless computing basada en Funciones. Al final, encontramos más aplicaciones de lo que inicialmente pensamos, permitiendo elaborar nuevas versiones de la solución en el futuro cercano.

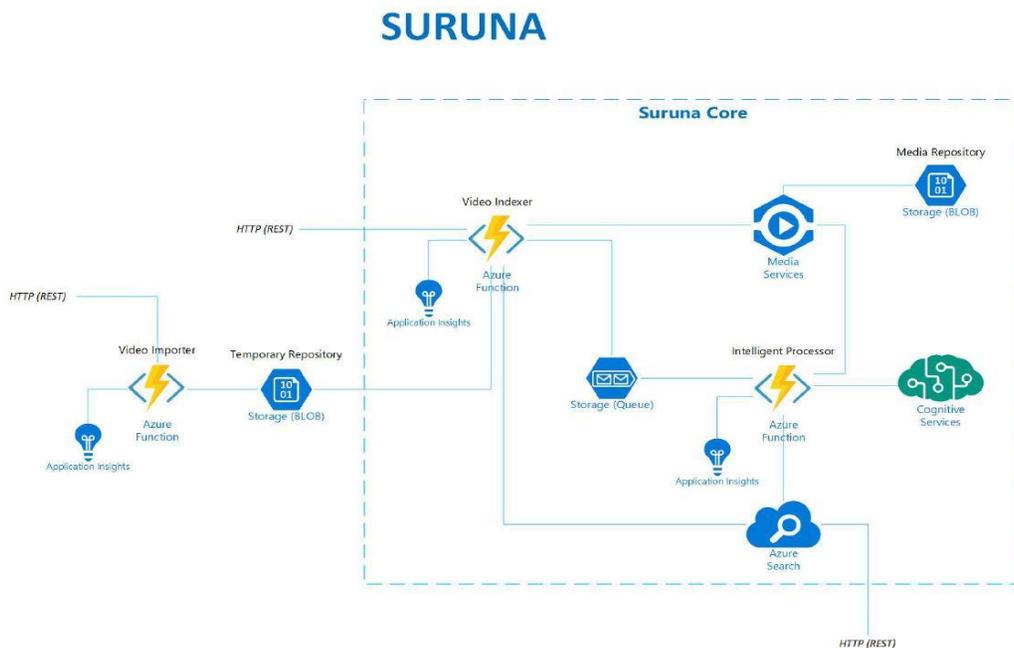


Figura 1. Diagrama de arquitectura de la solución
Fuente: Elaboración propia (2018)

diseño de la solución apunta a poder procesar masivamente una gran cantidad de videos sobre distintos temas.

Para la prueba caso “Lava Jato” Peru, los videos originales estaban almacenados previamente en una cuenta de Azure Blob Storage externa. Con el fin de procesar esos videos, desarrollamos un Azure Function del tipo "HTTP triggered function" para importar y mover los videos a otro Azure Blob Storage controlado por nosotros (Temporary Repository). Usamos esta definición JSON para invocar a la función como se muestra en la figura 4:

```
{
  "userName": "caso-odebrecht",
  "userKey": "<my user key>",
  "sourceContainerName": "myContainerName",
  "sourceSubdirName": "virtual/subdirs",
  "count": 0,
  "fileName": "",
  "filesExtension": ".mp4",
  "listFilesTxtContainerName": "myContainerName",
  "listFilesTxtFileName": "video-list.txt",
  "listFilesConnString": "<Blob Storage Connection String [Where you have .txt file with list of video names to proc",
  "sourceConnString": "<Blob Storage Connection String [Source of videos]>"
}
```

Figura 4. Configuración de parámetros de HTTP

Fuente: Elaboración propia (2018)

Si todos los parámetros enviados en el mensaje HTTP POST fueron correctos, entonces nosotros grabamos temporalmente todos los videos de la fuente original al Azure Blob Storage propio para iniciar con el flujo principal de procesamiento. Después de realizar una importación exitosa, es necesario un mensaje de confirmación para iniciar el proceso, y eso es realizado por otro llamado con la siguiente estructura:

```
{
  "userName": "caso-odebrecht",
  "userKey": "<my user key>",
  "indexLanguage": "esEs"
}
```

Figura 5. Formato del mensaje de confirmación

Fuente: Elaboración propia (2018)

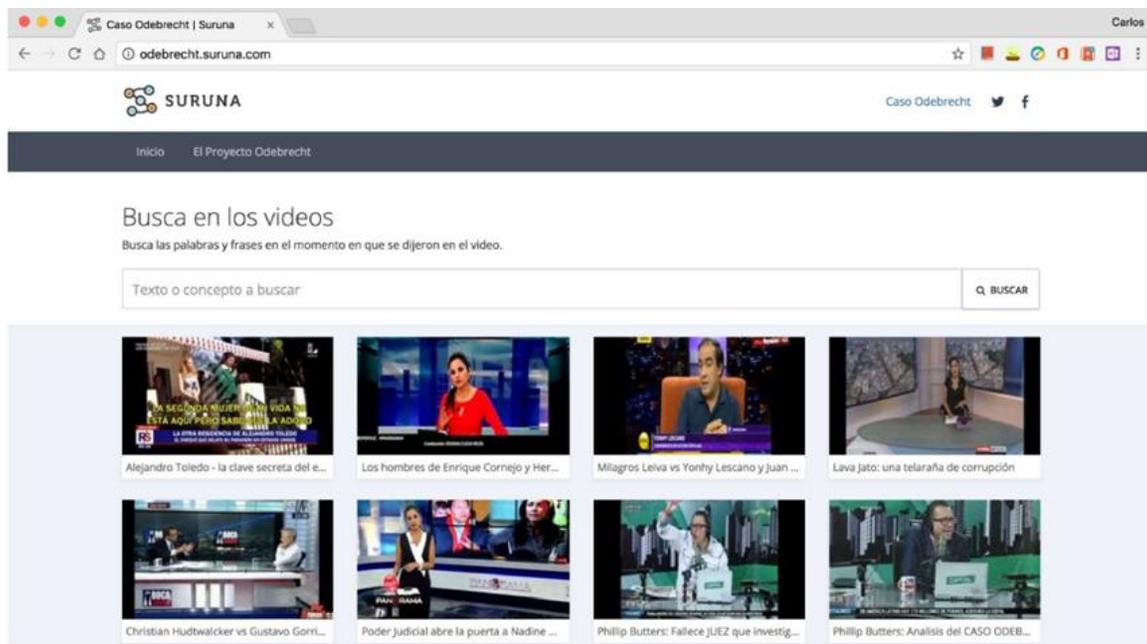
Aquí se especifica el idioma del proceso de indización. Este mensaje inicia una tarea para cada video que se encuentra en el almacenamiento temporal.

Posterior a ese procesamiento inicial, los textos obtenidos del audio son publicados en Azure Search Index (Microsoft, 2019a). Luego, se procesan estos textos generados en formato WebVTT y se consume el servicio Text Analytics de Microsoft Azure para extraer frases y palabras clave que se ingresarán también en el Azure Search Index. El API de Text Analytics devuelve una lista de textos que representan los temas clave del texto generado a partir del audio del video.

Es así como diseñamos una arquitectura capaz de satisfacer las necesidades futuras de potenciales clientes como se muestra en la figura 1, donde un cliente promedio tiene 3 mil videos a los que pueden acceder 1.5M veces aproximadamente por mes. Para este escenario específico, un diseño en máquinas virtuales (infraestructura como servicio) no sería eficiente. Además, el diseño permite la integración de otros servicios cognitivos según sea necesario.

3. Resultados

Esta solución se utilizó por primera vez, como prueba piloto, para procesar videos del caso de Odebrecht “Lava Jato”.



*Figura 6. Prueba de proceso de videos
Fuente: Elaboración propia (2018)*

Posterior a las pruebas piloto realizadas, los resultados fueron totalmente satisfactorios. De forma cuantitativa, se logró subir, indizar y publicar de forma automática 3 mil videos en

solo 3 horas de procesamiento, lo que nos da un promedio de mil videos por hora (aproximadamente 17 videos por minuto).

Se ha logrado automatizar 100% el procesamiento de videos que permite a los usuarios tener una plataforma digital donde puedan filtrar y buscar el contenido multimedia de forma eficiente, para las empresas de comunicaciones.

Además, la implementación de tecnologías de computación en la nube para el desarrollo de soluciones digitales servirá como medio de consulta tanto para las diferentes empresas de medios de comunicación y a las instituciones que desee soluciones para búsquedas y filtrados de textos y documentos que estén disponibles en la nube. Asimismo, un ahorro en los costos de implementación y un mejor retorno de inversión (Salesforce, 2018).

4. Conclusiones

1. La solución de gestión de videos permitió a los usuarios tener una plataforma digital donde puedan filtrar y buscar contenidos de multimedia de forma eficiente para las empresas de comunicaciones.
2. Los algoritmos de Servicios Cognitivos y Serverless de Microsoft Azure permitieron que el diseño elástico y modular de la arquitectura sea eficiente.
3. La implementación de módulos de importación – carga de videos es sencillo gracias al uso de servicios de cómputo en la nube del tipo “serverless”.
4. El uso de servicios de computación en la nube es más eficiente y fácil de administrar, si los comparamos con soluciones tradicionales.
5. Los algoritmos de Machine Learning son eficientes para recomendaciones de textos y videos.

Agradecimientos

Este artículo contó con la asesoría del Dr. Justo Solis, Decano de la Facultad de Ingeniería Electrónica e Informática en la Universidad Nacional Federico Villarreal. Además, se agradece al Mg. Ulises Román, quien ha colaborado con la revisión del contenido final y los alcances teóricos del artículo científico.

Referencias

- Microsoft. (2019a). *Azure Search*. Retrieved 2018, from Microsoft Azure:
<https://azure.microsoft.com/es-es/services/search/>
- Microsoft. (2019b). *Cognitive Services*. Recuperado el 2018, de Microsoft Azure:
<https://azure.microsoft.com/es-es/services/cognitive-services/>
- Redacción RPP. (29 de Noviembre de 2018). 6 claves para entender el caso Lava Jato. *RPP Noticias*. Recuperado el 2018, de <https://rpp.pe/mundo/latinoamerica/que-es-la-operacion-lava-jato-6-claves-para-entender-este-caso-noticia-943263>
- Salesforce. (2018). *12 Benefits of Cloud Computing*. Retrieved 2018, from Salesforce Hub:
<https://www.salesforce.com/hub/technology/benefits-of-cloud/>
- Shuler, R. (2002). *How does Internet Work?* (Pomeroy IT Solutions) Recuperado el 2018, de Stanford University: <https://web.stanford.edu/class/msande91si/www-spr04/readings/week1/InternetWhitepaper.htm>

