

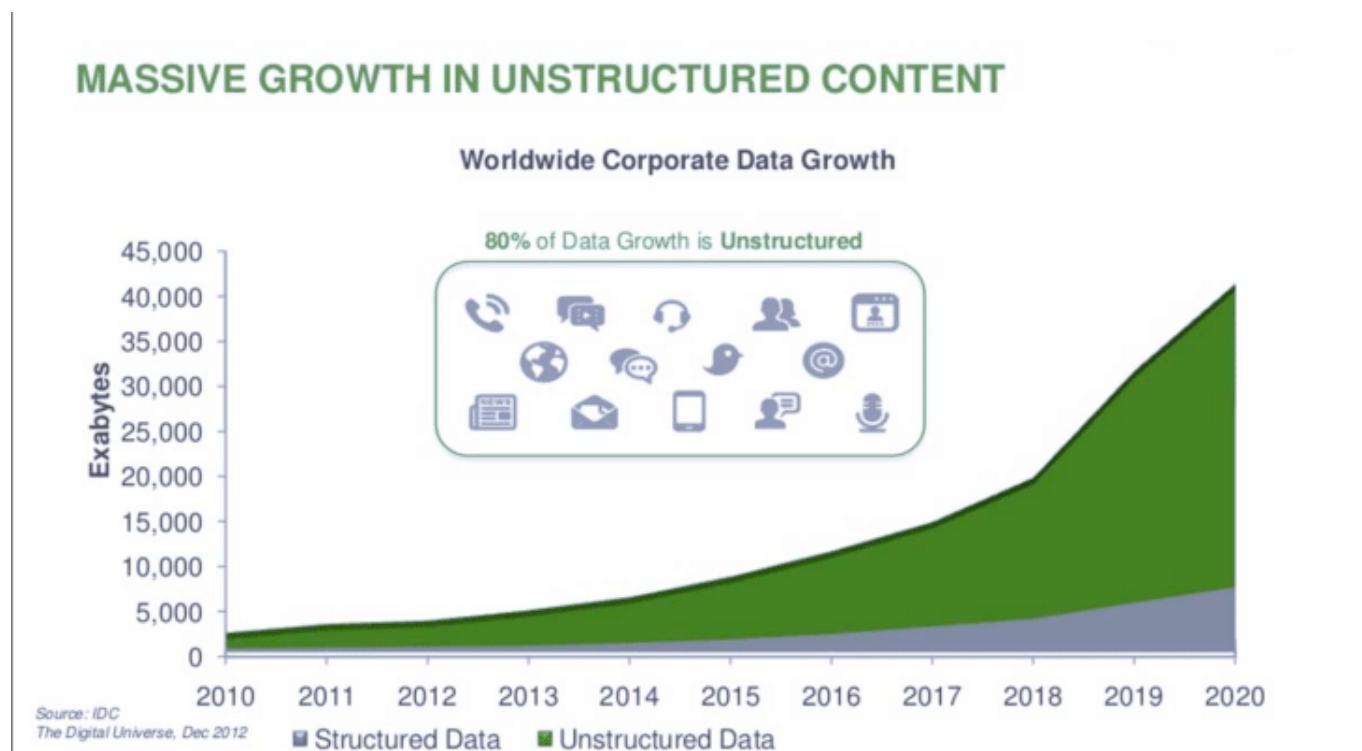
Artículo

[Dani Fibla](#) · 4 nov, 2022 · Lectura de 2 min

[Open Exchange](#)

Enriquece tus proyectos de analítica con PLN (Procesamiento de Lenguaje Natural)

Según la consultora IDC, el 80% de todos los datos producidos son NoSQL. Mira:



Hay documentos digitales y escaneados, textos online y offline, contenido BLOB (objeto binario grande) en SQL, imágenes, vídeos y audio. ¿Te imaginas una iniciativa de Analítica Corporativa sin todos estos datos para analizar y apoyar las decisiones?

En todo el mundo, muchos proyectos están utilizando tecnologías para transformar estos datos NoSQL en contenido de texto, para poder analizarlo. Fíjate:

1. Imágenes escaneadas e imágenes con texto extraído usando OCR (Google Tesseract es una buena opción);
2. Vídeos analizados con Visual Computing soportado por Machine Learning (OpenCV es una buena opción) y transformando los resultados en conjuntos de datos JSON o XML;
3. Contenido externo de Internet y Scraping en Redes Sociales usando Python y almacenando los resultados en contenido de texto.

Todo este contenido extraído se guarda como texto y podría ser analizado con motores de PLN, como InterSystems IRIS Text Analytics (iKnow).

Hay varias opciones para hacerlo:

1. Guardar en una tabla los datos de texto extraídos y crear un Dominio NLP (PLN en inglés) para esa tabla, mira:

Server 9843a27a7696Namespace IRISAPPSwitchUser _SYSTEMLicensed To InterSystems IRIS CommunityInstance IRIS

Analytics > Text Analytics > Domain Architect - (dc.ocr.OcrNLP)

Domain Architect

NewOpenSaveCompileBuildDelete

Model ElementsExpand AllCollapse All

OcrNLP (ID: 1)Element Type

▼ Domain Settings

▼ Metadata Fields

filenameSTRING

▼ Data Locations

OcrNLPTableTable:[dc_ocr.OcrTable]

▼ skiplists

▼ Matching

DetailsTools

Select an item on the left to view and edit its properties

NameOcrNLPTable

Batch Mode☒

Schemadc_ocr

Table NameOcrTable

ID FieldID

Group FieldID

Data Field

2. Usar una API NLP para enviar en tiempo real a NLP el texto extraído, así:

```
$SYSTEM.iKnow.IndexString("OcrNLP", pRequest.FileName, pRequest.Text, , 0, .src)
```

3. Guardar el texto extraído en archivos de texto y configurar la localización de los datos en una carpeta de archivos.

4. Crear un canal RSS para que NLP consuma el texto extraído.

Ahora, con tu NLP configurado puedes analizar los resultados:

quick brown dogExplore!

5

Top Conceptsfrequencydominance

regular151

os131

fraude61

area51

quick brown dog41

lazy fox41

born41

0541

consulta21

modeloiafraudeaqua21

Similar Entities

quick brown dog41

Related Conceptsrelatedproximity

lazy fox41

SourcesPathsCRCsCCs

1:TEMP:testocr.pngThe quick brown dog jumped over the lazy fox. ... The quick brown dog jumped over the lazy fox. ... The quick brown dog jumped over the lazy fox. ... The quick brown dog jumped over the lazy fox. ...

Sin esfuerzo, IRIS realizó el ranking de conceptos, agrupó entidades similares (cosas, hechos, nombres, sustantivos) y creó las relaciones entre entidades (conceptos), el CRC - Conceptos/Relaciones/Conceptos. Fue posible analizar la ruta para llegar a un concepto y se pueden usar colores para conocer características como sentimientos, negaciones y otras, incluyendo características modeladas en un diccionario customizado.

Para aprender y perfeccionar los resultados, IRIS NLP utiliza diccionarios, como: <https://github.com/intersystems-community/irisdemo-demo-twittersentiment...>

Finalmente, el análisis se puede consumir usando la API nativa de IRIS con Java, .NET, Python y Node.js. También se puede consumir como una API REST , mira: <https://docs.intersystems.com/irislatest/csp/docbook/Doc.View.cls?KEY=GI...>

Para ver todos los detalles de estos proyectos:

1. <https://openexchange.intersystems.com/package/Twitter-Sentiment-Analysis...>
2. <https://openexchange.intersystems.com/package/COVID-19-iKnow-Content-Nav...>
3. <https://openexchange.intersystems.com/package/OCR-Service>

[#Analítica](#) [#InterSystems](#) [IRIS](#)

[Ir a la aplicación en InterSystems Open Exchange](#)

URL de
fuente: <https://es.community.intersystems.com/post/enriquece-tus-proyectos-de-anal%C3%ADtica-con-pln-procesamiento-de-lenguaje-natural>