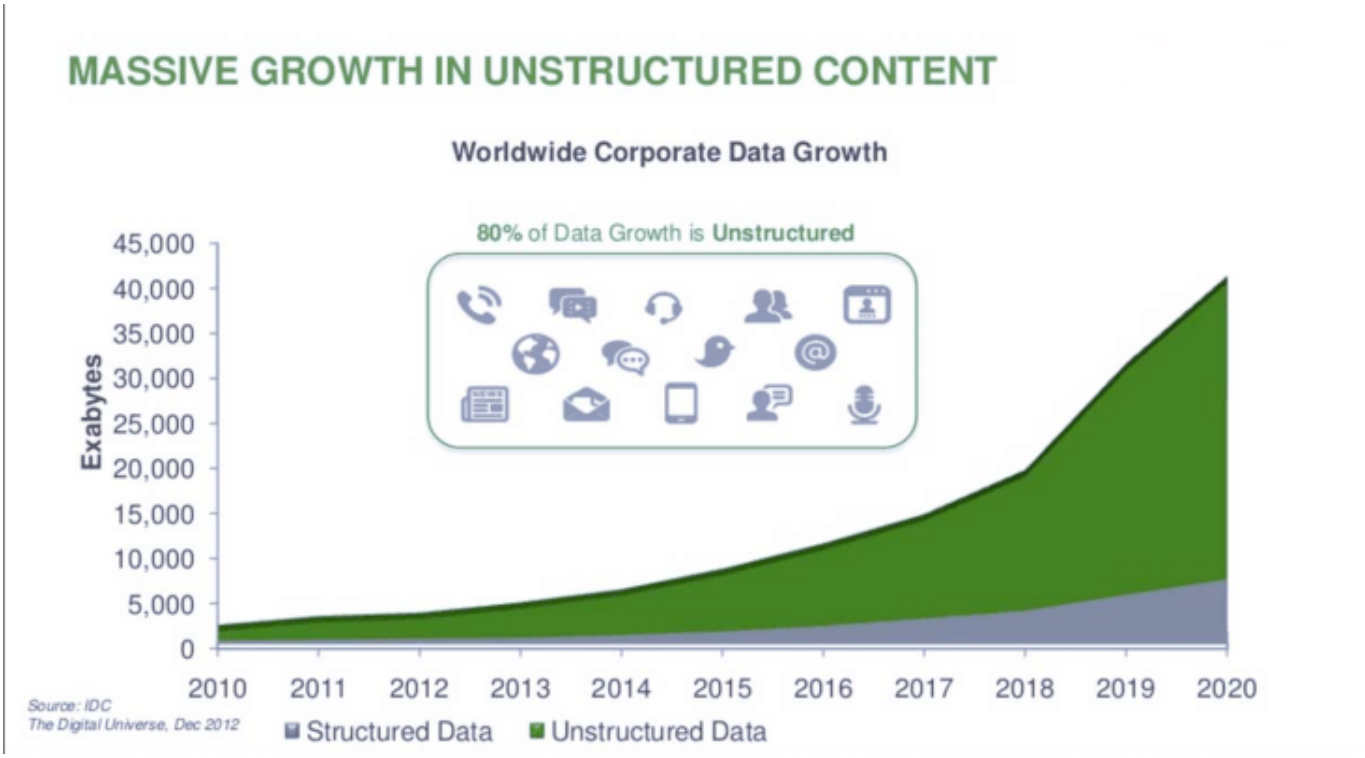
Article
Yuri Marx · Nov 20, 2020 2m read

Open Exchange

Enrich your analytics projects with NLP

According IDC, 80% of all data produced are NoSQL. See:



There are digital documents, scanned documents, online and offline texts, blob content into SQL, images, videos and audio. Imagine a Corporate Analytics initiative without all these data to analyze and support decisions?

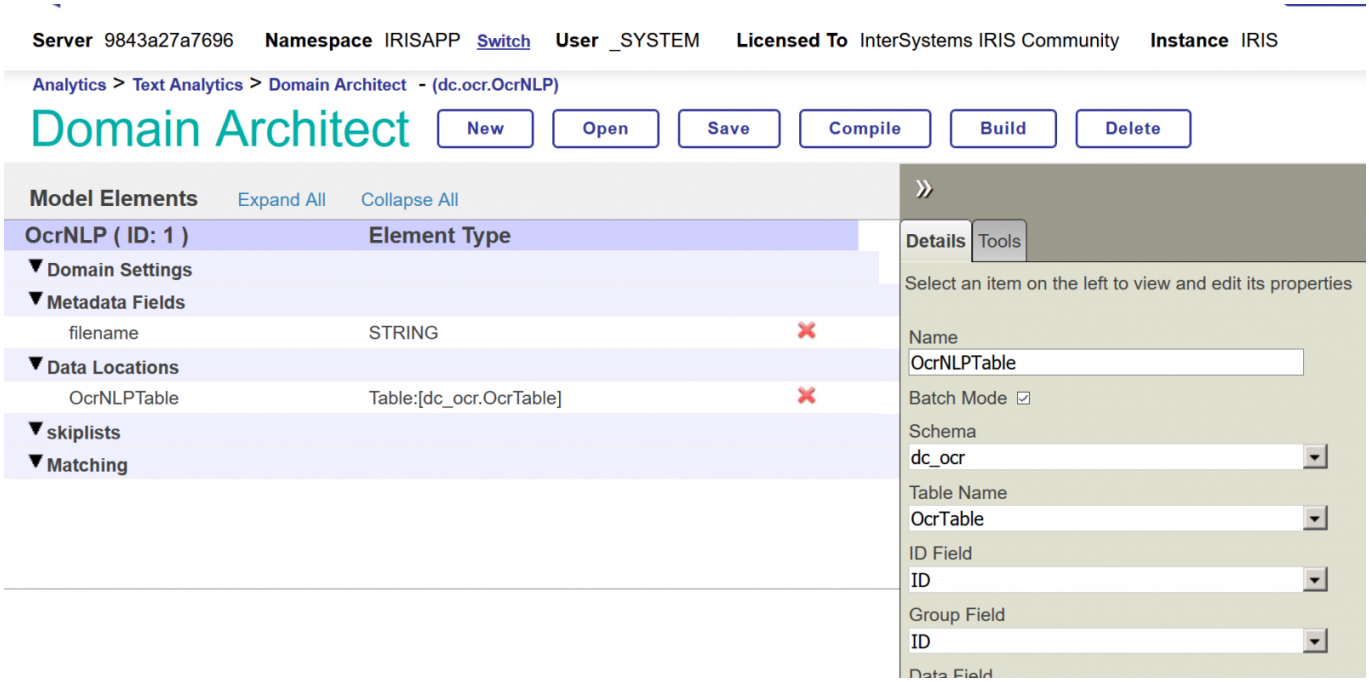
In all the world, many projects are using techonologies to transform these NoSQL data into textual content, to allows analyze it. See:

1. Scanned images and images with text extracted using OCR (Google Tesseract is a great option);
2. Videos analyzed with Visual Computing supported by Machine Learning (OpenCV is a good option) and transforming the results into JSON or XML dataset results;
3. External content from Internet and Social midia scraping using Python and storing results into textual content.

All these content extracted are stored into text, and could be analyzed with NLP engines, like InterSystems IRIS Text Analytics (iKnow).

There are some options to do this:

1. Store textual data extracted to a table and create a NLP Domain to this table, see:



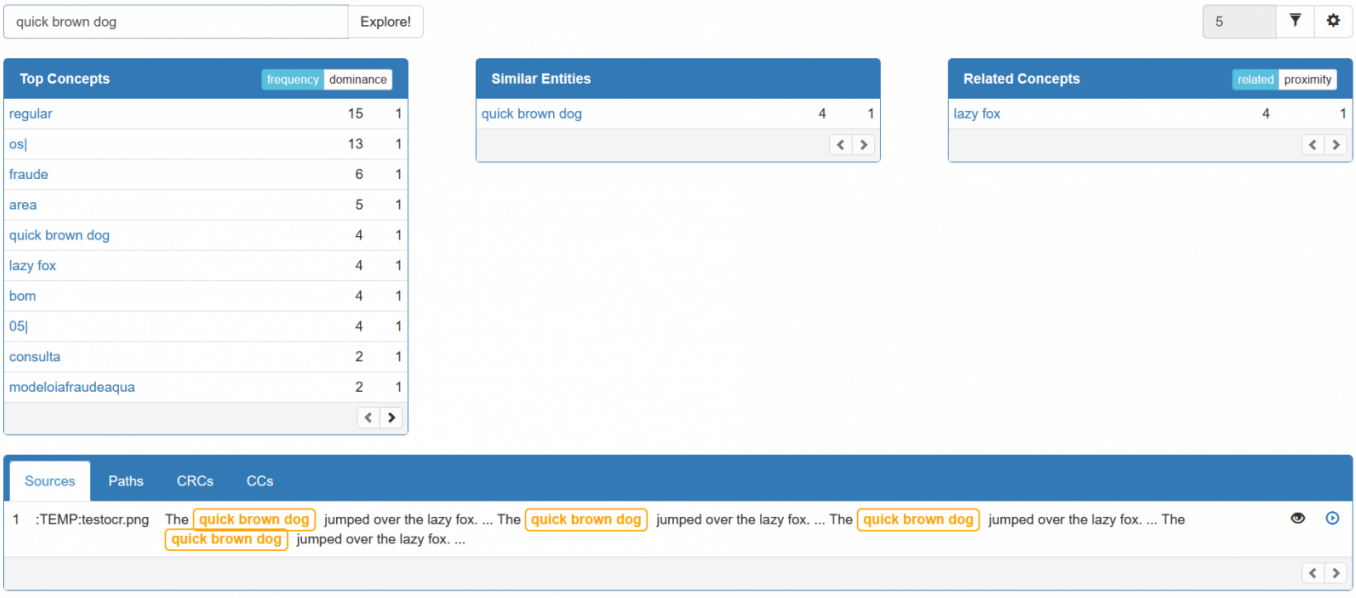
2. Use NLP API to send extracted text to NLP in realtime, see:

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

3. Save extracted text to text files and set data location to files folder.

4. Create RSS channel to NLP consume the text extracted.

Now, with your NLP configured you can analyze the results, see:



With no effort, IRIS did the ranking of concepts, cluster similiar entities (things, facts, names, substantives) and created the relationships between entities (concepts), the CRC - Concepts/Relations/Concepts. It was possible analyze the path to reach a concept and could be used colors to know features like sentiments, negations and other features, including features modeled into a custom dictionary.

To training and refine results, IRIS NLP use dictionaries, like it: https://github.com/intersystems-community/irisdemo-demo-twittersentiment...

Finally, the analysis may be consumed using IRIS native API with Java, .NET, Python and Node.js. Can be consumed as REST API too, see: https://docs.intersystems.com/irislatest/csp/docbook/Doc.View.cls?KEY=GI...

To see all details see these projects:

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

#Analytics #InterSystems IRIS
Check the related application on InterSystems Open Exchange