

Customer Success Story

Boehringer Ingelheim

The leading pharmaceutical company Boehringer Ingelheim requested a scientific information tracking service. In research intensive businesses, scientific publications are fundamental assets which need to be leveraged. A semantic data portal on internal and external publishing activities with flexible analytic features was built.

The challenge

The requested information and content assets are distributed internally and externally. Analytic functionalities must be applied to structured and unstructured data.

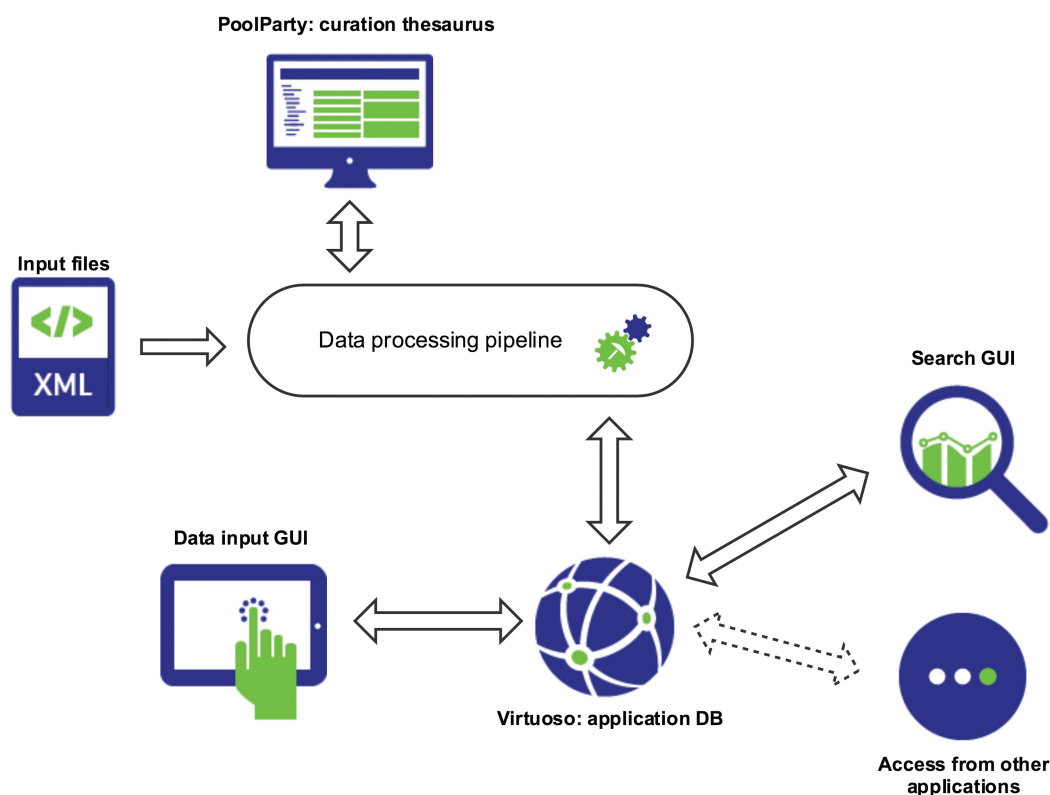
The solution

The tracking application includes a faceted search and visualisations that show for example how different locations of Boehringer Ingelheim compare in terms of quality (impact factors of the publications) and quantity (number of publications).

The architecture is based on Semantic Web technologies. An open platform was created that is accessible for other applications. In the future, the data can be easily linked in a standard way across multiple platforms and be reused for other purposes.

The results

The head of science of the pharmaceutical company has a reliable application where he can track the publication activity of the companies global workforce. The single point of access substantially reduces his controlling tasks.



Project insights

How to develop semantic search portals



STEP 1: DATA PROCESSING

The data input to the system are publication information like title, abstract, authors, affiliations, citation information, ... This data is processed in a pipeline that produces a RDF representation that gets written to a RDF triple store (a Virtuoso Universal Server in this case).



STEP 2: DATA NORMALIZATION

As the data is heterogenous and the same concepts can be expressed in different ways, the data must be normalized. This is the case for example with:

- ✓ authors (e.g. is "Smith, J" and "Smith, JH" the same person?)
- ✓ journals
- ✓ locations (can be given in different languages as "Vienna" vs. "Wien")
- ✓ misspellings



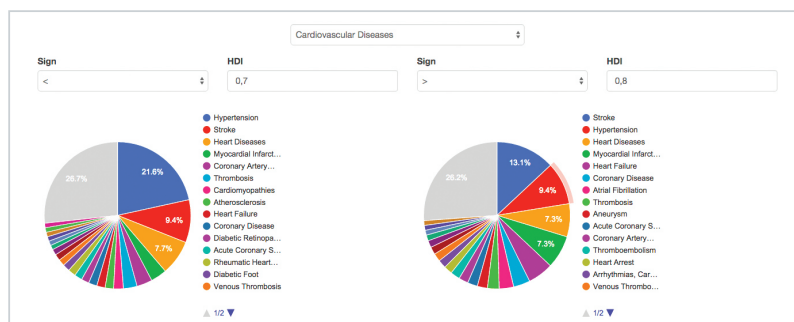
STEP 3: CONNECTING THE DOTS

The actual database of the system is a Virtuoso triple store and the publication monitoring and search application is built directly on top of this store. The application connects via the standard SPARQL endpoint and includes a faceted search and data and trend visualisation components.

TAKE A LOOK

http://integrator.poolparty.biz/report_medicine/

As the scientific tracking service of Boehringer Ingelheim includes confidential data, it can't be shown publicly. However, the PoolParty Semantic Integrator Demo exemplifies how structured and unstructured data can be combined.



REACH OUT TO THE PROJECT TEAM

Do you want to know more? Contact us!



ALEXANDAR KAPISODA

Project Lead - Boehringer Ingelheim

Mail: aleksandar.kapisoda@boehringer-ingelheim.com



CHRISTIAN BLASCHKE

Senior Consultant - PoolParty Suite

Mail: c.blaschke@semantic-web.at