The NLP Use Case Database Project

A bottom-up qualitative NLP success failure database

Key Information

Project Manager: Birthe Nielsen

birthe.nielsen@pistoiaalliance.org

SC members:

Etzard Stolte, Roche (Project champion)

Bikalpa Neupane, Takeda

Vera Hazelwood, Astra Zeneca

Sponsors/partners

Pistoia Alliance: Lowering barriers to R&D innovation

The Pistoia Alliance is a global, not-for-profit alliance of life science companies, vendors, publishers, and academic groups that work together to lower barriers to innovation in R&D.

Our members collaborate as equals on open projects that generate significant value for the worldwide life sciences community.

Background

Pharma companies apply Natural Language Processing (NLP) methods in hopes of automation and insight generation, and many companies are investing in NLP initaitatives to stay ahead of the competition. Although NLP algorithms have matured quite a bit during the past years, practical value for most NLP pilots tends to be poor, and very few NLP-driven projects are seen through to production. In addition, the issues facing the Pharma companies are :

- Hard to determine upfront, if a use-case will work
- "Successful" pipelines do not work for slightly different data
- Manual curation, training sets are exceedingly work-intensive
- Success seems very use case specific
- Expensive to find talent that actually delivers the quality we need

Value Propostion & Deliverables

The Pistoia Alliance therefore chartered a project to

- Share best-practices – get an idea what works & what does not work
- Build a bottom-up qualitative NLP database of use cases; incl. failures
- Agree on annotations of NLP use case methods and success/failure criteria
- Gain collaborative insight into why NLP use cases may fail or succeed within an industry — wide view

Where the database stands today:

The database contain a range of the following NLP types:

- Named-entity recognition (NER)
- Concept Extraction (taxonomy extraction, concept tagging, document classification, semantic search)
- Multi-Label Text-Classification
- Unsupervised topic modeling
- Extractive auto-summarization
- Similarity Search
- Chatbots/Virtual Assistants/Q&A service
- Text Data Augmentation
- Text translation
- Topic discovery
- NLP Document prediction
- Natural Language text to Python code generation
- Relationship Extraction for Knowledge Graphs
- Quality Analytics
- Patent Annotation

Database Content

Use Case
- Deployment type
- Project Level
- Company
- Language
- etc

Data
- Description
- Size
- Issues
- Preparation
- FTE involved/team
- etc

Algorithm
- Summary
- Version
- Liberay
- etc

Outcome
- Learnings
- Success rating

Discussions
- Effort/Benefit ratio
- Non-NLP approach
- Support used
- Lessons learned

Oulook
- Build vs Buy recommendations
- Other applications of technology
- etc

Conclusion

By creating a reference database, Pharma members will be able to narrow down successful use case scenarios, resulting in less experimentation, and more successes. This type of knowledge is of value to share in a pre-competitive manner among Pistoia Alliance members. The success of any Natural Language Processing solution will be in its ability to deliver quantifiable Return on Investment. Ultimately, the goal is for you to spend less time doing manual work and ensure that you make the most of your text, to get the answers you need.

Join the project:

Get in touch to join the project – use cases wanted:
NLPdatabase@pistoiaalliance.org or
birthe.nielsen@pistoiaalliance.org

www.pistoiaalliance.org

@pistoiaalliance