Semantic Search and Analysis: Exploiting word and sentence level embeddings

- Use of distributional semantics for addressing vocabulary mismatch issues in search
- Semantic representation from raw data in an unsupervised fashion using machine learning
- Learning data associations and patterns given initial query and focused taxonomy

**Industry Benefits**

- Address query-document mismatch vocabulary issues in search for diverse users
- Learn semantic insights from high volume of raw big data collections
- Categorization of social media streams and feeds
- Adapt search and categorization solutions to new domains

**Use Cases**

- Search solutions for specialist domains, e.g. legal, medical, physical sciences
- Semantic Analysis for social media and feed classification, e.g. Tweets, RSS, news feeds
- Personalization and adaptation of search systems by mining and analyzing user search behaviour
- Local archive and intranet search systems for enterprise and other organisations

**Industry Challenge:**

**Content Search and Categorization**

- Addressing vocabulary mismatch issues between search queries and information represented in documents
- Developing and adapting search solutions for diverse users
- Handling rapid increase in potentially relevant information in social media and online information feed updates
- Categorizing and clustering of daily business related information to predict risk and market shift
- Expanding coverage of small in-house curated lexicons for specialist data classification

**Our Solution:**

**Word and Sentence level embeddings**

Semantic representation of words and sentences using machine learning based embedding methods to build search solutions and feed categorization.

- Semantic expansion of words in taxonomies
- Improved categorization using expanded taxonomies
- Learn word and sentence level representation of queries and documents to capture semantics effectively
- Compare queries and documents using word and sentence level distributed representations

**Summary:** Provide enhanced search solutions combining traditional word level search and categorization models with semantic representation based on effective and robust vector-embedding based models.

Piyush Arora, Murhaf Hossari, Diab team and Prof. Gareth Jones

To learn more about innovative ADAPT technologies, contact: collaboration@adaptcentre.ie