Referat Joseph Birkner (Montag, den 15.05.2017)

Ranking With Neural Network Derived Document Vectors

First of all we he was talking about the project „**IROM**“ which stands for „**Intelligent Recommendation of Massive Open Online Courses“**
The vision was about Ubiquitous Vertical Search. So basically every Information Need can be satisfied instantly with a vertical search engine. It means that a search is within a specific doman. Though solution for specific domain may be applicable to other domains at well at the end.
Domain-specific search solutions focus on one area of knowledge, creating customized search experiences, that because of the domain's limited corpus and clear relationships between concepts, provide extremely relevant results for searchers.

In the domain-specific setting one can combine the tf-idf approach implemented via an inverse index with semantic approaches of semantic headers and semantic skeletons. Instead of most frequent keywords, a set of entities is extracted from a portion of text to be matched against a potential question. This allows much more flexibility due to real-time reasoning capabilities while matching questions and answers in the form of semantic headers.

Any general search engine would be indexing all the pages and searches in breadth first manner to collect documents. Whereas, the spidering in domain specific search engines is more efficient which is through searching a small subset of documents by focussing on particular set. The spidering that can be accomplished using reinforcement learning framework which allows optimal behaviour, which is three times more efficient than breadth-first search as per experimental results.

Next he was talking about the Motiviation for encoding of documents.
We need efficient document representations to instantaneously rank recommended courses based on student need.
So the goal here is to develop an envolvong system that can generate good document representations for effective and efficient ranking, which are enabling fast (constant time). This is for efficiency and intelligent ranking, which stands for effectiveness.

Furthermore he was talking about Word2vec, which was created by a team of researchers led by Tomas Mikolov at Google. The algorithm has been subsequently analysed and explained by other researchers.Embedding vectors created using the Word2vec algorithm have many advantages compared to earlier algorithms like Latent Semantic Analysis.

Also he mentioned Doc2Vec, which modifies the word2vec algorithm to unsupervised learning of continuous representations for larger blocks of text, such as sentences, paragraphs or entire documents.

And at the end he told us about his tasks which is creating his prototype of his own to demonstrate his work.