Paper Summaries

Swoogle: a search and metadata engine for the semantic web

https://dl.acm.org/doi/pdf/10.1145/1031171.1031289

Authors presented Swoogle, a crawler-based indexing and retrieval system for the Semantic Web. Metadata is extracted for each discovered document, and relations between documents is computed. Relevant documents can be found by using either character N-Gram or URIrefs as keywords to find relevant documents and to compute the similarity among a set of documents. Ontology rank, a measure of the importance of a Semantic Web document is computed. Ranks of Semantic Web Documents are computed using a rational random surfing model.

Context-Aware Intent Identification in Email Conversations

https://www.microsoft.com/en-us/research/uploads/prod/2019/05/Wang_SIGIR19.pdf

The paper focuses on sentence-level intent identification in email conversations and presents that incorporating more context, such as the full message body and other metadata improves the performance of the intent identification models. Enterprise email dataset: Avocado corpus was used. Four email intent categories were included: information exchange, task management, scheduling and planning, and social communication. Experiments were done with several models for leveraging context including both classical machine learning(TF-IDF features and SVM/LR models) and deep learning approaches(Dynamic-Context Recurrent Neural Network (DCRNN) model with bi-directional GRU as sentence encoder.)

Structuring Wikipedia Articles with Section Recommendations

https://arxiv.org/pdf/1804.05995.pdf

The present paper proposes several approaches for the section recommendation for Wikipedia articles. The recommendations are generated by sourcing sections from articles that are similar to the input article. Several ways are explored for defining similarity like based on topic modeling, collaborative filtering, and Wikipedia's category system. Both automatic and human evaluation

approaches are used for assessing the performance of the recommendation system. Category-based approach works best and achieves precision@10 of about 80% in the human evaluation.

The Many Kinds of Creepware Used for Interpersonal Attacks http://damonmccoy.com/papers/Creepware_SP.pdf

A larger landscape of apps, called creepware exist that are unstudied and used for interpersonal attacks. Authors developed a new algorithm, CreepRank to help surface previously unknown examples of creepware, by using the principle of guilt by association, which is then characterized through a combination of quantitative and qualitative methods. As a result of the authors' work, Google Play Store has removed hundreds of apps for policy violations.

Mapping the Underground: Supervised Discovery of Cybercrime Supply Chains http://damonmccoy.com/papers/ecrime2019.pdf

Generating the supply chains performed in a cybercrime requires time consuming manual effort. Authors proposed a method that leverages machine learning and graph-based analysis to efficiently extract supply chains from cybercrime forums. The analysis of the supply chains demonstrates underlying connections between products and services that are potentially useful understanding and undermining the illicit activity of these forums. The approach was evaluated using two popular forums: Antichat (Russian) and Hack Forums (English). The supply chain is built by analyzing the posts and replies and a chronological record of which products the users of a forum bought and sold. The classifiers categorize the posts and replies, and build an interaction graph to build supply chain.

Relevance Assessment: Are Judges Exchangeable and Does it Matter?

http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.156.6386&rep=rep1&type=pdf

Authors have investigated that the people making relevance judgements for a reusable IR test collection, to what extent are exchangeable. Three classes of judge are considered: 1) gold standard judges: topic originators and experts in a particular information seeking task; 2) silver standard judges: task experts but do not create topics; 3) bronze standard judges: who do not define topics and are not experts in the task. The analysis shows low levels of agreement in relevance judgements between the three groups. Experiments were performed to determine if it is sufficient to invalidate the use of a test collection for measuring system performance when silver or bronze standard judges create the relevance assessments. Both system scores and system rankings were found to be mostly consistent. Test collections were not found completely robust to changes of judge when the judges vary widely in task and topic expertise. Bronze standard judges may not be able to substitute for topic and task experts, and gold standard judges are preferred. The measure, infNDCG estimates a well-known performance score from a small number of judgements.