"Ku semantycznym miarom zawartości informacyjnej - integracja wiedzy z grafu linków i semistrukturalnej informacji tekstowej".

Mieczysław Kłopotek
Instytut Podstaw Informatyki PAN, Warszawa, mieczyslaw.klopotek@ipipan.waw.pl

Celem projektu jest prowadzenie badań w obszarach:

- learning to rank –nauczenie się z interakcji z człowiekiem sposobu rangowania dokumentów w odpowiedzi na zapytanie do wyszukiwarki internetowej;
- sentiment analysis –stwierdzenie zabarwienia emocjonalnego dokumentu (co implikuje subiektywność itp.);
- ekstrakcja hierarchii pojęć z dokumentów –automatyczne tworzenie tezaurusów pojęć podobnych, nadrzędnych i podrzędnych w celu wspomagania wyszukiwania faktów i dokumentów;
- wykrywanie spamerskich dokumentów, serwisów itd.

W ramach każdego z tych zagadnień celem będzie:

- ewaluacja dotychczasowych modeli;
- opracowanie polepszonych metod i modeli;
- stworzenie zasobów wspomagających realizację danego zadania;•stworzenie metod i zasobów ewaluacji metod;
- integracja wybranych metod/modeli z systemem semantycznej wyszukiwarki internetowej;
- poszukiwanie efektów synergicznych między poszczególnymi obszarami oraz ich eksploatacja w celach (1) usprawnienia działania wyszukiwarki –jako element badań stosowanych (2) pogłębienia rozumienia konceptu wartości informacji jako element badań podstawowych.

Rozumienie wartości informacji z punktu widzenia ludzkiej percepcji jest nie tylko z teoretycznego lecz także z inżynierskiego punktu widzenia jednym z najistotniejszych konceptów tej części informatyki, która zajmuje się przetwarzaniem informacji w Internecie, ponieważ dotychczasowa interpretacja zawartości informacyjnej, bazująca na entropii Shannona, wydaje się być całkowicie nieadekwatna.
W szczególności palącym problemem jest takie dodefiniowanie informacji, które odpowiadało jej ludzkiemu rozumieniu i pozwalałoby na automatyczną ocenę informacji w sposób zadawalający człowieka.
Do elementów takiej oceny należą między innymi badania z takich obszarów jak np. wyżej wymienione learning to rank, sentyment analysis, concept hierarchy extraction czy spam detection.

Zalecana literatura

[1] Zhou, Lina, 2007: „Ontology learning: state of the art and open issues”, Information Technology and Management”, 2007/09 pp. 241- 252 UR - https://doi.org/10.1007/s10799-007-0019-5DO - 10.1007/s10799-007-0019-5

[2] Kaity, Mohammed and Balakrishnan, Vimala, 2019: "An automatic non-English sentiment lexicon builder using unannotated corpus", The Journal of Supercomputing. 2019/75, pp. 2243—2268

[3] Roffo, Giorgio, 2017: “Ranking to Learn and Learning to Rank: On the Role of Ranking in Pattern Recognition Applications.”. arXiv http://arxiv.org/abs/1706.05933