

A SURVEY OF COLLABORATIVE WEB SEARCH

Through Collaboration among Search Engine Users
to More Relevant Results
[a position paper]



Charles University in Prague
Czech Republic

Pavel Surynek



Motivations

○ Motivation for (web) **search**

- growing amount of data
 - textual, visual
 - automated collecting
 - at www level, but also at level of companies



○ Motivation for **collaboration**

- increase accessibility of data
 - user is unsure how to express what she/he is searching
 - help from other users of search service
 - automate the exchange of information among users
 - find people searching similar items



Web Search and Recommender Systems

○ Recommender systems

- a system for recommending items
 - suitable for recommending **music, movies, books, ...**
 - collects what items were preferred by individual users
 - identify your preference
 - find user with similar preference
 - recommend you an item you may also like

○ Recent development

- items within social network
 - profile similarity → similar preference
 - Facebook's Graph Search
 - search engines build implicit profiles

Recommending high performance camera

		Nikon D7000	Canon 7D	Nikon D800	Sony α77
					
Pavel					
Anna					
Josef					

Overview of Web Search (1)

○ Software engineering aspect

- crawling
 - web pages/documents are visited and content is mined
 - new web pages/documents discovered through links
- indexing
 - words and phrases are associated with their occurrences
 - enables fast word/phrase searching
 - resource demanding process
 - external data structures, distribution, synchronization
- interacting
 - queries from users are processed
 - collecting user's behavior



Overview of Web Search (2)

○ Computational intelligence aspect

- result ordering
 - variants of page-rank algorithm
 - random surfers – randomly follow a link
 - pages where many random surfers gather are important
- understanding the search query
 - natural language processing
 - language dependent stemming
 - search of semantically related terms
- proprietary know-how
 - difficulty for making research



Overview of Recommender Systems

- **Item-based** filtering
 - similar items to already preferred ones are recommended
- **Collaborative** filtering
 - preferred items gathered from many users
 - able to recommend novel items
- **Matrix factorization** techniques
 - user preferences collected in a sparse matrix
 - rows correspond to **users**, columns correspond to **items** ... **A**
 - find **small** matrices X and Y , such that $X \times Y \approx A$
 - latent features are discovered



Collaborative Web Search

- Extend **collaborative filtering** to web search
 - use semantic terms instead of items
 - ontologies needed to detect semantically equivalent terms

- **Challenges**

- diverse users



- similar preferences in one area of interest does not imply similarity in others
 - example:** users like similar movies but are different in spending free time
 - need to introduce ontologies on areas of interest

- **Problematic evaluation**

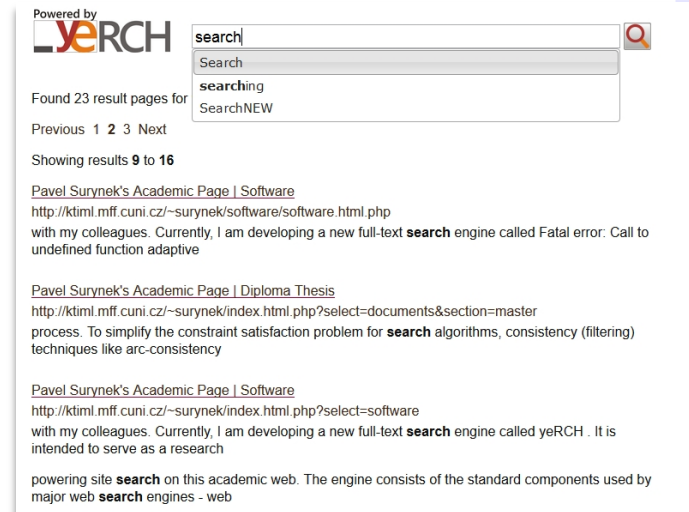
- new benchmarks need to be suggested

yeRCH: Our Experimental Search Engine



<http://ktiml.mff.cuni.cz/~surynek/yerch/>

- Designed for **small web sites**
 - web shop/intranet
 - academic site
- Current state
 - standard functionality
 - recommender module collects data
 - challenge – increase traffic
- Observation
 - users make few related searches



Written in C++, 150K lines of code

Conclusions

- Position paper addressing **web search**
 - overview of *web search*
 - software engineering aspect
 - computational intelligence aspect
 - overview of *recommender systems*
 - matrix factorization
 - discovering latent features
- **Collaborative** web search
 - apply recommender techniques in web search
 - items are expressed implicitly
 - difficulty with diverse users