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 \rightarrow similar preference
 - Facebook's Graph Search
 - search engines build implicit profiles



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×Y≈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

Lecret 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

	search
	Search
F 100 V 6	searching
Found 23 result pages for	SearchNEW
Previous 1 2 3 Next	
Showing results 9 to 16	
Pavel Surynek's Academic	: Page Software
http://ktiml.mff.cuni.cz/~su	rynek/software/software.html.php
with my colleagues. Currel	ntiy, I am developing a new full-text search engine called Fatal error. Call
Pavel Surynek's Academic	nty, i am developing a new iuli-lexi search engine called Fatal error. Call e : <u>Page Diploma Thesis</u>
Pavel Surynek's Academic http://ktiml.mff.cuni.cz/~sur	nty, i am developing a new iuli-lexi search engine called Fatal error. Call e : <u>Page Diploma Thesis</u> rynek/index.html.php?select=documents§ion=master
with the contemposities of the contempositie	Huy, Fam developing a new full-text search engine called Fatal error. Call e <u>Page Diploma Thesis</u> rynek/index.html.php?select=documents§ion=master instraint satisfaction problem for search algorithms, consistency (filtering) ency
Pavel Surynek's Academic http://ktiml.mff.cuni.cz/~su process. To simplify the co techniques like arc-consist Pavel Surynek's Academic	tiny, i am developing a new iuli-lext search engine called Fatal error. Call e : <u>Page Diploma Thesis</u> ynek/index.html.php?select=documents§ion=master nstraint satisfaction problem for search algorithms, consistency (filtering) ency : <u>Page Software</u>
An information of the second s	tiny, Lam developing a new full-lext search engine called Fatal error. Call e : Page Diploma Thesis rynek/index.html.php?select=documents§ion=master instraint satisfaction problem for search algorithms, consistency (filtering) ency : Page Software rynek/index.html.php?select=software
Pavel Surynek's Academic http://ktml.mft.curi.cz/-su process. To simplify the co techniques like arc-consist Pavel Surynek's Academic http://ktiml.mft.curi.cz/-su with my colleagues. Currei intended to serve as a ress	tiny, I am developing a new full-lext search engine called Fatal error. Call e <u>Page Diploma Thesis</u> ynek/index.html.php?select=documents§ion=master instraint satisfaction problem for search algorithms, consistency (filtering) ency <u>Page Software</u> ynek/index.html.php?select=software ntly, I am developing a new full-text search engine called yeRCH. It is earch

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