Learning to Rank - the next big thing in search?

Learning to Rank is a way to apply machine learning techniques to improve the ranking of search engine results and is a current focus of organisations including Microsoft Research, here in Flax's home town of Cambridge.

The order, or rank, of search engine results has traditionally been performed by search engine software using algorithms such as the vector space model or the probabilistic BM25. Learning to Rank takes user feedback (for example in the form of clickthrough data recorded on a website) as training data, combines this with features of the result documents (for example, document length) and feeds this into a learning algorithm. The search engine thus 'learns' how best to re-order the results, improving relevance for future users of the system.

The open source search engine Xapian, which Flax has used in many projects for clients including the Financial Times and Durrants, recently joined the Google Summer of Code (GSoC) programme, where Google funds students to work on open source projects. A Learning to Rank feature for Xapian was one of the projects selected by the Xapian team for a student to work on out of a record number of submissions to the project. We hope that the features developed during the project will become available in the software later this year, which will allow us to apply these innovative features to search applications.

Find out more about Learning to Rank at

http://en.wikipedia.org/wiki/Learning_to_rank

and about the Xapian GSoC projects at

http://trac.xapian.org/wiki/GSoC2011

Events

Charlie Hull from Flax will be speaking at the Guardian's Government Procurement show, SmartGov Live, on 14th/15th June this month. Charlie is also helping to organise a major search event this autumn, Enterprise Search Europe, to be held by the Information Today organisation in London on 24th/25th October. Flax also runs the popular Cambridge Enterprise Search Meetups; informal gatherings for search professionals held every few months, the next being on 22nd June upstairs at the Boathouse Pub on Chesterton Road, Cambridge.

www.smartgovlive.com/

www.enterprisesearcheurope.com/2011/

www.searchmeetups.com/

Flax offers best practises

Our open source search experts have been carrying out short 'best practises' consulting engagements for various organisations including Cambridge University Press and Boliven, the patent search engine. These engagements have focussed on how to get the best out of existing search engine applications based on Apache Lucene/Solr, including how to improve indexing throughput, search speed and accuracy. Our staff are backed by expert training and support supplied by our US partner Lucid Imagination, the commercial company behind Apache Lucene/Solr and will shortly become certified Apache/Solr engineers under Lucid's new certification programme. If you have an existing search application you'd like us to analyse, do get in touch.

About Flax

Flax is highly active in the information retrieval market with international clients from sectors including academia, public relations, e-commerce, government and private businesses. Flax's clients include The Financial Times, Accenture, the Newspaper Licensing Authority (NLA), The University of Cambridge and Mydeco. Flax delivers a cutting-edge enterprise search solution, using the power of open source software to drive down costs and provide world beating search performance with no software licence fees. Flax is an authorized partner of Lucid Imagination, the commercial company behind Lucene & Solr.