WWW 2010 Tutorial: Applications of Open Search Tools

Rosie Jones
Yahoo! Labs
4 Cambridge Center
Cambridge, MA
jonesr@yahoo-inc.com

Ted Drake
Yahoo! Inc
701 First Ave
Sunnyvale, CA 94089
tdrake@yahoo-inc.com

ABSTRACT

It costs about 300M dollars to just build a search system that scales to the web. Web services which open up web search results to the public allow academics, developers, and entrepreneurs to achieve instant web search parity for free and enable them to focus on building their additional secret sauce on top to create even grander relevant services. For example, a social network could leverage open search and their data about users to personalize web search. Additionally, one of the best ways to gather data about web search behavior is to build your own search system. Prototype IR and web search systems based on open search can be used to gather user interaction data and test the applicability of research ideas. Open Source tools like Lucene and Nutch and open search services like from major search engines can greatly help developers implement these types of systems quickly, allowing for fast production and evaluation. We will give detailed overviews of the popular open search tools, showcasing examples of search and IR algorithms and systems implemented using them, as well as discussing how evaluation can be performed.

Categories and Subject Descriptors

H.3.5 [**Information Search and Retrieval**]: Online Information Services—*Web-based services*; H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing; H.3.4 [**Information Storage and Retrieval**]: Systems and Software

General Terms

Algorithms, Experimentation, Performance

Keywords

open search tools, open source search

1. BIOGRAPHIES

Rosie Jones is a Senior Research Scientist in Information Retrieval at Yahoo! Labs. She is an active participant in the IR community, serving as Senior PC member for SIGIR since 2007. Her research interests include information retrieval, web search and natural language processing.

Rosie co-taught a tutorial on open search at SIGIR in 2009. She co-taught the tutorial on web search and data mining at SIGIR 2008. In 2005 she co-organized the SIGIR workshop on lexical cohesion and information retrieval, and in 2003 she co-organized the ICML workshop on the Continuum from Labeled to Unlabeled Data in Machine Learning and Data Mining. She obtained her PhD from the Language Technologies Institute at Carnegie Mellon University.

Ted Drake is a front end engineer for Yahoo!. He has worked on multiple properties, including Finance, Tech, Answers, as well as platform technology, accessibility, and browser testing.

Outside of Yahoo!, Ted has built several food related search engines on open search platforms. He was recently spotlighted on the Yahoo! Developer Network for his regional food search engine: http://insiderfood.com. Ted has spoken regularly at international Yahoo! and web technology conferences, including presenting "the Future of Vertical Search Engines" at the WWW2009 Conference in Madrid. Ted has also had a previous career as a professor of Photography at Palomar College, and has a Bachelor of Fine Arts from San Diego State University.