



Report on **INEX 2009**

Shlomo Geva Jaap Kamps Ralf Schenkel Andrew Trotman

INEX 2009 Workshop was held in Brisbane, December 6-10, 2009

Motivation: Focused Retrieval

- Common experience on the Web
 - ★ Find a perfect result, and the snippet shows part of the answer
 - ★ Click on the result, and...
 - ★ ... a long page loads
 - ★ and I need to read/search through the page to find it back

- Can't we do better than that?
 - ★ Make use of widely available document structure → retrieve parts of documents directly
 - ★ Make use structured hints in the queries → powerful queries

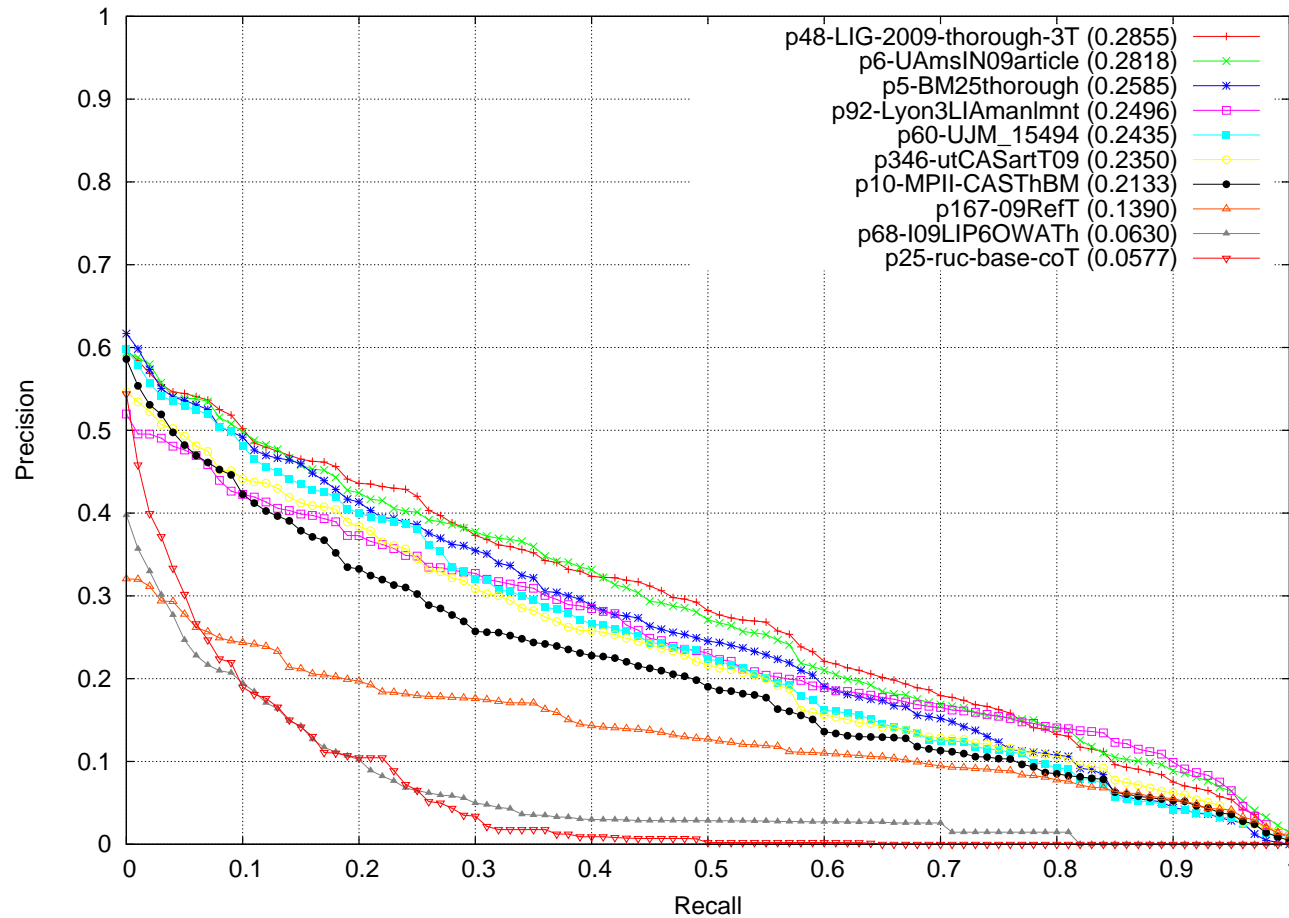
INEX 2009 Tracks (1)

- Featured eight tracks
 - ★ **Ad hoc Track** The main track of INEX 2009 will investigate the effectiveness of XML-IR and Passage Retrieval for three ad hoc retrieval tasks (Focused, Relevant in Context, Best in Context).
 - ★ **Book Track** Investigating techniques to support users in reading, searching, and navigating full texts of digitized books.
 - ★ **Efficiency Track** Investigating the trade-off between effectiveness and efficiency of XML ranked retrieval approaches on real data and real queries.
 - ★ **Entity Ranking Track** Investigating entity retrieval rather than text retrieval: 1) Entity Ranking, 2) Entity List Completion.

INEX 2009 Tracks (2)

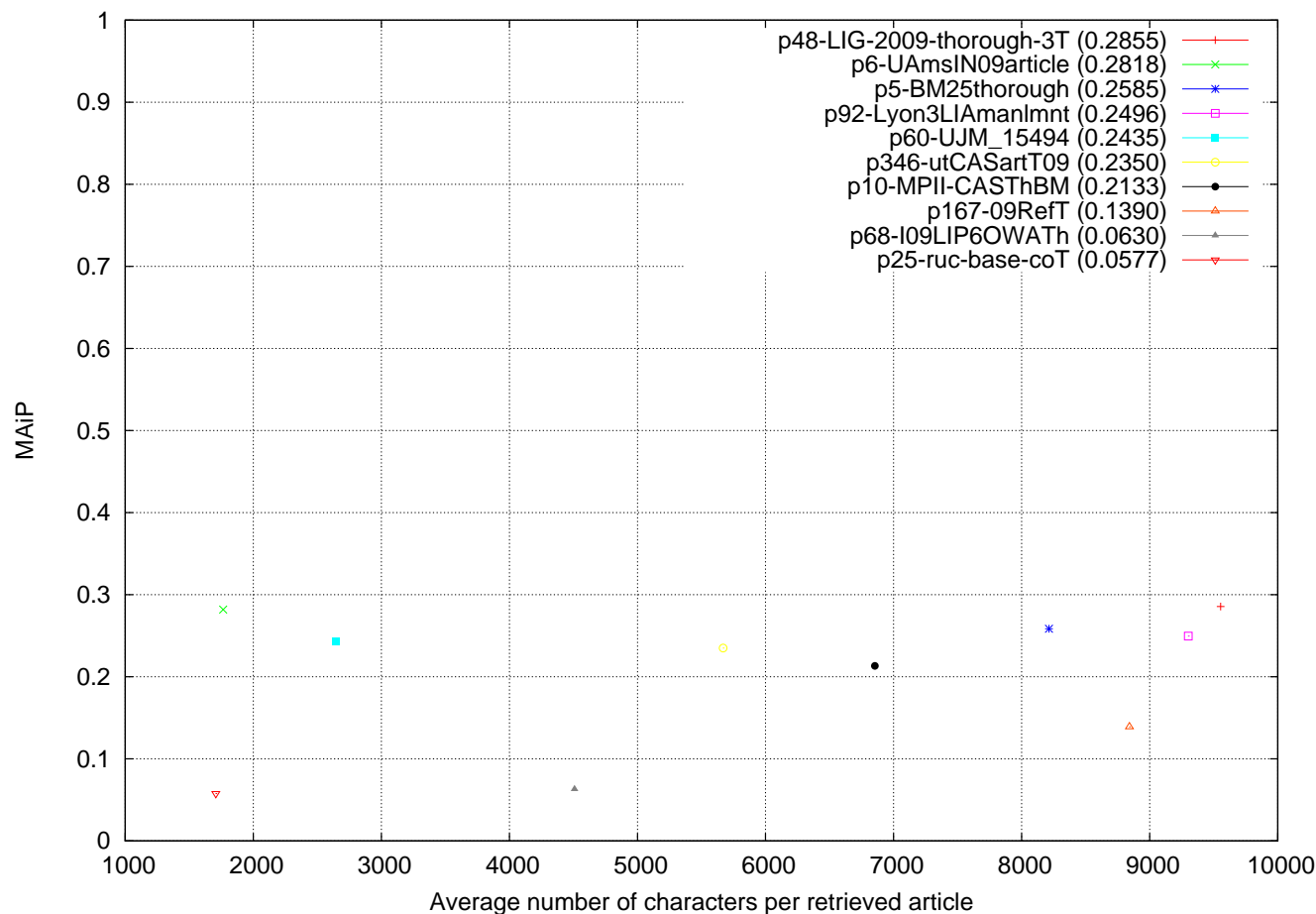
- Continuing:
 - ★ [Link-the-Wiki Track](#) Investigating link discovery between Wikipedia documents, both at the file level and at the element level.
 - ★ [XML-Mining Track](#) Investigating structured document mining, especially the classification and clustering of semi-structured documents.
 - ★ [Interactive Track \(iTrack\)](#) Investigating the behavior of users when interacting with XML documents.
 - ★ [Question Answering \(QA@INEX\) Track](#) Investigating technology for accessing semi-structured data can be used to address natural language questions.

Some highlights (Ad Hoc)



- Top 10 runs for the Thorough task.

Score versus length retrieved elements (Ad Hoc)



- Top 10 runs for the Thorough task.

Viewing Focused Results as Snippets

XML Find: GNU GPL

http://berk.science.uva.nl:8080/xmlfind/result.php?srcQuery=GNU+GPL

Getting Started Latest Headlines news mail search uva blog amsterdam vefrit Wikiii

Search: GNU GPL Search

xmlfind

Display results 1 to 10 of 24 (Lucene took 0.335 sec. to return results, XML Find took 2.591 sec. to display results)

1 2 3 next

Open Source Software Development: An Overview
Ming-Wei Wu, Ying-Dar Lin

- article[1]
 - body[1]
 - sec[2] (OPEN SOURCE BACKGROUND)

... for work on the **GNU** Project (<http://www.gnu.org/gnu/thegnuproject.html>). **GNU** is a recursive acronym for "**GNU**'s Not Unix" and a homophone for "new." The **GNU** Project seeks to develop ... principal author of the **GNU** C Compiler (GCC), **GNU** symbolic debugger (GDB), **GNU** ... in essence, copyrights with **GPL** regulations. Open source software, essentially ...
 - sec[4] (LICENSING MODELS)
 - p[5]

... code falls under the **GPL** model, nVidia had to remove all **GPL** code, then re-release the ... Software Systems, makers of **GPL**-regulated software, and repackaged its ... work, thereby skirting the **GPL** regulation.
 - p[6]

... multilicensing models to avoid **GPL** violations. For example, Sun's ... adapts three licensing models: **GPL**, **LGPL**, and **SISSL**. Ordinary users who can fulfill **GPL** regulations can use StarOffice under **GPL**. Proprietary developers and companies can use **LGPL** or **SISSL**, which both ...
 - sec[5] (BUSINESS MODELS)

... software company cannot fulfill **GPL** regulations. Thus, a private ... its reputation. For example, **GNU** has a reputation for ...

Legal Implications of Operating Systems
Larry Graham

- article[1]
 - body[1]
 - sec[4] (DEVELOPING LINUX)
 - p[5]

The **GNU** software, which comprises about ... General Public License. The **GPL** takes advantage of copyright ... remains free. Under the **GPL**, anyone may freely copy, ... it. More important, the **GPL** requires that any subsequent ... the terms of the **GPL**. Accordingly, when others distribute ... are also free. The **GPL** itself may not

Clips from the Computists' Weekly
Kenneth I. Laws

- article[1]
 - body[1]
 - sec[4] (INTELLECTUAL PROPERTY)
 - brief[1]
 - p[1]

...

Done

INEX 2010

- Ongoing still
 - ★ **Adhoc Track** similar tasks but focus on user effort (measures from Tampere) and viewing focused retrieval as 'snippet like' summarization.
 - ★ New **Data Centric Track** using highly structured data based on IMDB.
 - ★ New **Web Service Discovery Track** on huge collection of services with WSDL descriptions
 - ★ **Interactive Track** works on large set of Amazon data (including formal descriptions and user-generated data).
- Most tracks continue.

Questions?

- INEX Web site <http://inex.otago.ac.nz/>
- INEX 2010 Workshop is December 13-15 in the Netherlands!