Scalable, Flexible and Generic Instant Overview Search
Pavlos Fafalios, Ioannis Kitsos and Yannis Tzitzikas

Presentation by: Pavlos Fafalios
(fafalios@csd.uoc.gr)

FORTH-ICS
Information Systems Laboratory

University of Crete
Computer Science Department
What is Instant Search?
What is Instant Overview Search (IOS)?

Clustering

Apache Tomcat - Welcome!
Download, documentation and tutorials for the straight-forward servlet container and Web server. Apache Tomcat was the Servlet Container reference implementation and ...
http://tomcat.apache.org/

Tomcat | Define Tomcat at Dictionary.com
noun 1. a male cat. 2. Slang . a woman-chaser. Dictionary.com"s Mobile Apps www.dictionary.com/mobile Ad-Free, Offline Content, Audio Pronunciation and More! Take ...
http://dictionary.reference.com/browse/tomcat

Apache Tomcat - Apache Tomcat 6 Downloads
Tomcat 6 Downloads: Welcome to the Tomcat 6.x download page. This page provides download links for obtaining the latest version of Tomcat 6.0.x, as well as links to ...
http://tomcat.apache.org/download-60.cgi
What is Instant Overview Search (IOS)?

Meta-data based groupings
What is Instant Overview Search (IOS)?

Entities

Organization (174 entities)
- Tomcat (12)
- Apache (5)
- Apache Software Foundation (4)
- Java (4)
- MotoMco Ltd (2)
- Rodent (2)
- Rodent Baits & Stations (2)
- Java Community (2)
- Windows (2)
- Manfrotto (2)

Location (47 entities)
- US (3)
- United States (3)
- Connecticut (2)
- Florida (2)
- Historic Park (2)
- Islands (2)
- London (2)
- California (3)
- Texas (2)
- City (1)
What is Instant Overview Search (IOS)?

Apache Tomcat - Welcome!
Download, documentation and tutorials for the straight-forward servlet container and Web server. Apache Tomcat was the Servlet Container reference implementation and ...
http://tomcat.apache.org/

Tomcat | Define Tomcat at Dictionary.com
noun 1. a male cat. 2. Slang . a woman-chaser. Dictionary.com's Mobile Apps www.dictionary.com/mobile Ad-Free, Offline Content, Audio Pronunciation and More! Take ... 
http://dictionary.reference.com/browse/tomcat

Apache Tomcat - Apache Tomcat 6 Downloads
Tomcat 6 Downloads: Welcome to the Tomcat 6.x download page. This page provides download links for obtaining the latest version of Tomcat 6.0.x, as well as links to ...
http://tomcat.apache.org/download-60.cgi
What is Instant Overview Search (IOS)?

User has typed only 2 characters

Results of selected clusters: reset

JavaServer Pages - Apache Tomcat
Tomcat is a free, open-source implementation of Java Servlet and JavaServer Pages technologies developed under the Jakarta project at the Apache Software Foundation.
http://java.sun.com/products/jsp/tomcat/

About Tomcat - IBM - United States
This topic provides information about the Apache Software Foundation
Jakarta Tomcat servlet engine.

The Jakarta Site
The Jakarta Downloads
BSF; Commons; DB; Excalibur; Gump; HiveMind; HttpComponents; James; JCS; JMeter; Logging; Lucene; Maven; POI; Portals; Struts; Taglibs; Tapestry; Tomcat; Turbine; Velocity; Watchdog
http://jakarta.apache.org/site/downloads/index.html

The Jakarta Site - The Apache Jakarta™ Project -- Java Related ...
What is Instant Overview Search (IOS)?

User has typed only 2 characters

Results of selected clusters: reset

JavaServer Pages - Apache Tomcat
Tomcat is a free, open-source implementation of Java Servlet and
JavaServer Pages technologies developed under the Jakarta project at the
Apache Software Foundation.
http://java.sun.com/products/jsp/tomcat/

About Tomcat - IBM - United States
This topic provides information about the Apache Software Foundation
Jakarta Tomcat servlet engine.
    /rzaetomcat.htm

The Jakarta Site Jakarta Downloads
BSF; Commons; DB; Excalibur; Gump; HiveMind; hhttpComponents; James;
JCS; JMeter; Logging; Lucene; Maven; POI; Portals; Struts; Taglibs;
Tapestry; Tomcat; Turbine; Velocity; Watchdog
http://jakarta.apache.org/site/downloads/index.html

The Jakarta Site - The Apache Jakarta™ Project -- Java Related ...
Key Challenges

- **Efficiency and Scalability**
  - *Offer real-time interaction*
  - *Exploit available main memory and disk*

- **Flexibility**
  - *Reduce user’s effort*

- **Generality**
  - *Ability to accommodate various kinds of supplementary information*

*Using very modest hardware!*

*(server running on an ordinary personal computer)*
Our approach: *Partitioned Trie-based Index*

**Hosted in main memory** (based on requests and space)

- **sub-trie1**: int: file, int: bytes to skip, int: bytes to read
- **sub-trie2**: int: file, int: bytes to skip, int: bytes to read
- **sub-trie3**: int: file, int: bytes to skip, int: bytes to read

**Precomputed information** (always in hard disk)

**Average Retrieval Time** $\approx 135\text{ms}$

*Experiments* over a server running on a modest personal computer, with a synthetic query log of 1 million distinct queries and synthetic precomputed information of 1 Terabyte
Our approach: *Hybrid Caching Scheme*

- **The problem:**
  - A large number of users start typing queries at the same time.
  - The system requires loading multiple subtries, i.e. the appropriate subtrie for each user’s keystroke.
  - The system may get overloaded!

- **Solution:** **Adopt a caching scheme**
  - Keep in memory a number of subtries
  - Options: full static cache, full dynamic cache, hybrid cache

- A hybrid caching scheme (70% static and 30% dynamic) offers*:
  - **80% better throughput**
  - **25% speedup** for queries that lie in the index

*Experiment: Synthetic query log of 1 million distinct queries and synthetic precomputed information of 1 Terabyte*
Our approach: *Tolerate Typos and Different Word Order*

Typo-tolerant instant search
Our approach: *Tolerate Typos and Different Word Order*

Word order independent instant search
Current IOS Prototypes

1. Meta-search engine (over Google) offering instant **clustering** of the **snippets** of the top-100 results.

2. Standalone web search engine offering instant **meta-data based groupings** of the search space.

3. Meta-search engine (over Google) offering instant **entity mining** of the **contents** of the top-50 results.

4. Meta-search engine (over Google) offering both instant **entity mining** of the **contents** and instant **clustering** of the **snippets** of the top results.
Conclusion

- **IOS**: A more powerful *search-as-you* type functionality.
  - Predicts our search and shows results and *supplementary information* before finish typing

- The overall **effectiveness** of IOS, apart from its efficiency, depends on the **quality** of the precomputed information.
  - However, IOS can be applied over **ANY autocompletion, ranking, clustering, entity mining, etc.**, algorithm
Further Research

• Measure the effect of IOS to user satisfaction
  – We already gather user feedback by keeping a detailed log

• How exactly users exploit the precomputed information that appears instantly
  – This requires very fast eye-tracking equipment
Running prototypes:

http://www.ics.forth.gr/isl/ios

Thank you!