

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?



The screenshot shows a search engine interface with a search bar containing the text "wikipedia". To the right of the search bar is a microphone icon and a close button (X). Below the search bar is a dropdown menu with the following suggestions: "wikipedia", "wiki", "winrar", "wiz khalifa", and "wilfred". To the right of the search bar is a blue search button with a magnifying glass icon and the text "Advanced search". Below the search bar, the text "About 1,010,000,000 results (0.07 seconds)" is displayed. The first search result is for "Wikipedia" with a magnifying glass icon, the URL "www.wikipedia.org/ - Cached", and the description "Wikipedia, the free encyclopedia that anyone can edit." The second search result is for "Wikipedia, the free encyclopedia" with a magnifying glass icon, the URL "en.wikipedia.org/ - Cached", and the description "A free encyclopedia built collaboratively using wiki software. (Creative Commons Attribution-ShareAlike License). English - Deaths in 2011 - Game of Thrones - Adele". The third search result is for "Wikipedia - Wikipedia, the free encyclopedia" with a magnifying glass icon, the URL "en.wikipedia.org/wiki/Wikipedia - Cached", and the description "Wikipedia began as a complementary project for Nupedia, a free online English-language encyclopedia project whose articles were written by experts and ...".



What is Instant Overview Search (IOS)?

The screenshot shows a search interface with a search bar containing 'tomcat' and a 'Search' button. Below the search bar, there are suggestions: 'tomcat', 'tomcat apache', 'toyota', and 'toyota cars'. The main search results are displayed below, including links to 'Apache Tomcat - Welcome!', 'Tomcat | Define Tomcat at Dictionary.com', and 'Apache Tomcat - Apache Tomcat 6 Downloads'. On the right side, a search results panel shows a hierarchical cluster of results for 'tomcat(50)'. The cluster is highlighted with a red box. A callout box labeled 'Clustering' points to this cluster.

to

Search

tomcat

tomcat apache

toyota

toyota cars

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>

tomcat(50)

- ▶ apache(14)
- ⊕ free(9)
- ▶ download(4)
- ⊕ home(4)
- ▶ definition(5)
- ▶ definitive guide(3)
- ▶ java(9)
- ▶ page(5)
- ▶ cat(4)
- ⊕ tom(4)
- ▶ jason brittain senior(2)
- ▶ jakarta(4)
- ▶ dictionary(3)
- ▶ definitions(2)
- ▶ definition tomcat(2)

Clustering



What is Instant Overview Search (IOS)?

The screenshot shows a search interface with a search bar containing 'to' and a 'Search' button. Below the search bar, a list of suggestions is shown: 'tomcat', 'tomcat apache', 'toyota', and 'toyota cars'. The 'tomcat' suggestion is highlighted. Below the suggestions, three search results are displayed, each with a title, a brief description, and a URL. The sidebar on the right, highlighted with a red border, shows three meta-data based groupings: 'By date', 'By filetype', and 'By language'. The 'By date' grouping shows results for the years 2011 (33), 2010 (17), 2009 (15), 2008 (11), and 2007 (9). The 'By filetype' grouping shows results for application/msword (16), application/pdf (16), application/ms-powerpoint (3), and text/html (70). The 'By language' grouping shows results for Any (UTF-8) (27) and English (4).

to

Search

tomcat

tomcat apache

toyota

toyota cars

[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>

By date

- + 2011 (33)
- + 2010 (17)
- + 2009 (15)
- + 2008 (11)
- + 2007 (9)

By filetype

- ▶ application/msword (16)
- ▶ application/pdf (16)
- ▶ application/ms-powerpoint (3)
- ▶ text/html (70)

By language

- ▶ Any (UTF-8) (27)
- ▶ English (4)

Meta-data based groupings



What is Instant Overview Search (IOS)?

to

Search

tomcat

tomcat apache

toyota

toyota cars

[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>

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)

[show all](#)

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)

Entities



What is Instant Overview Search (IOS)?

The image shows a search interface with a search bar containing the text 'to' and a 'Search' button. Below the search bar, a list of suggestions is shown: 'tomcat', 'tomcat apache', 'toyota', and 'toyota cars'. The 'tomcat' suggestion is highlighted in blue. To the right of the suggestions, a red box highlights a blurred area of search results. A white diagonal banner with the text 'precomputed aggregated information' is overlaid on this red box. To the right of the red box, there is a white box with a large black question mark.

to Search

tomcat

tomcat apache

toyota

toyota cars

[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>

precomputed aggregated information

?



What is Instant Overview Search (IOS)?

to

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.
<http://publib.boulder.ibm.com/infocenter/iserics/v5r3/topic/rzaie/rzaietomcat.htm>

The **Jakarta** Site **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 ...

- tomcat(50)
 - ▶ apache(14)
 - ▶ free(9)
 - ▶ download(4)
 - ▶ home(4)
 - ▶ definition(5)
 - ▶ definitive guide(3)
 - ▶ java(9)
 - ▶ page(5)
 - ▶ cat(4)
 - ▶ tom(4)
 - ▶ iason brittain senior(2)
 - ▶ **jakarta(4)**
 - ▶ dictionary(3)
 - ▶ definitions(2)
 - ▶ definition tomcat(2)




What is Instant Overview Search (IOS)?

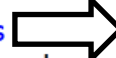
to

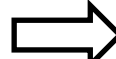
User has typed only 2 characters


Results of selected clusters: [reset](#)

- tomcat(50)
 - ▶ apache(14)
 - ⊕ free(9)
 - download(4)
 - ⊕ home(4)
 - ▶ definition(5)
 - ▶ definitive guide(3)
 - java(9)
 - ▶ page(5)
 - ▶ cat(4)
 - ⊕ tom(4)
 - ▶ on brittain senior(2)
 - ▶ jakarta(4)
 - ▶ dictionary(3)
 - ▶ definitions(2)

JavaServer Pages - Apache Tomcat  Position: 11 (2nd page)
Tomcat is a free, open-source implementation of server-side 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  Position: 16 (2nd page)
This topic provides information about the Apache Software Foundation Jakarta Tomcat servlet engine.
<http://publib.boulder.ibm.com/infocenter/iserics/v5r3/topic/rzaie/rzaietomcat.htm>

The Jakarta Site Jakarta Downloads  Position: 22 (3rd page)
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 ...  Position: 38 (4th page)



Key Challenges

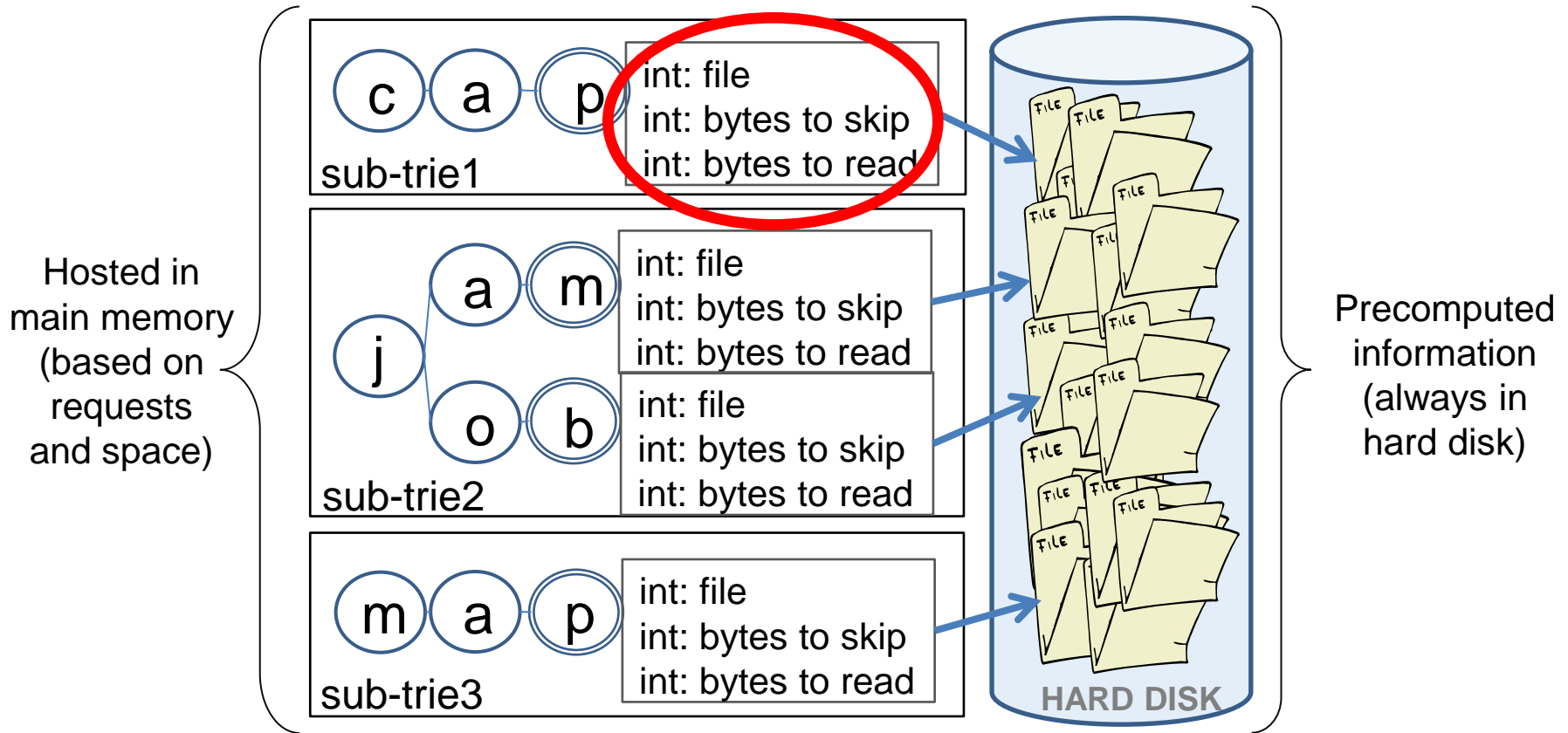
- Efficiency and Scalability
 - *Offer real-time interaction*
 - *Exploit available main memory and disk*
- Flexibility
 - *Reduce user's effort*
- Generality
 - *Ability to accommodate various kinds of supplementary information*

Using very modest hardware!

(server running on an ordinary personal computer)



Our approach: *Partitioned Trie-based Index*



Average Retrieval Time \approx **135ms**

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*

The screenshot shows a search interface with a logo on the left that says "clustering ios" with a magnifying glass over the "i". A search bar contains the text "Haeti e" with a red arrow pointing to it. Below the search bar is a dropdown menu of suggestions, with "haiti earthquake effects" highlighted in blue and enclosed in a red box. To the right of the search bar is a "search" button. Below the search bar, there is a list of search results, including "haiti earthquake effects(100)" and several sub-folders like "haitian(11)", "impact(14)", "january(12)", "nature(12)", "health(8)", "news(5)", "damage(21)", "devastating(6)", "effect(6)", and "disaster(6)".

clustering ios

Haeti e

search

haiti education
haiti economy
haiti earthquake
haiti earthquake effects
embassy of haiti
environment of haiti

haiti earthquake effects(100)

- ▶ haitian(11)
- ▶ impact(14)
- ▶ january(12)
- ▶ nature(12)
- ▶ health(8)
- ▶ news(5)
- ▶ damage(21)
- ▶ devastating(6)
- ▶ effect(6)
- ▶ disaster(6)

ke in Haiti

ts of earthquake in Haiti. The year 2010
terrible effects in Haiti that took up to 100000

www.twinside.org.sg/title2/gtrends/gtrends282.htm

Describe the effects of the Haiti earthquake

Typo-tolerant
instant search



Our approach: *Tolerate Typos and Different Word Order*



The screenshot shows a search interface with a search bar containing the text "Haeti e". Below the search bar is a dropdown menu with the following suggestions:

- haiti education
- haiti economy
- haiti earthquake
- haiti earthquake effects
- embassy of haiti**
- environment of haiti**

The search button is labeled "search". To the right of the search bar is a search results list:

- haiti earthquake effects(100)
 - haitian(11)
 - impact(14)
 - january(12)
 - nature(12)
 - health(8)
 - news(5)
 - damage(21)
 - devastating(6)
 - effect(6)
 - disaster(6)

Below the search bar, there is a snippet of text: "e in Haiti" and "s of earthquake in Haiti. The year 2010" and "terrible effects in Haiti that took up to 100000". At the bottom left, there is a URL: "www.twinside.org.sg/title2/gtrends/gtrends282.htm".

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!

