Exploratory search

Research Prototypes for Information Systems Laboratory
FORTH-ICS
Contact person: Yannis Tzitzikas
Last modification: August 1, 2016

OUTLINE

- Context
- A few words about Exploratory Search
- An overview of research prototypes from ISL
  - m1: faceted search using static and dynamic metadata
  - m2: instant overview search
  - m3: entity mining over documents
  - m4: configurability of entity mining
  - m5: application in professional search
  - m6: application in an research infrastructure
  - m7: preference-enriched faceted search
  - m8: top-k semantic graphs

Context

- Users should enjoy their data

EXPLORATORY SEARCH

Wikipedia:

"**Exploratory search** is a specialization of information exploration which represents the **activities carried out by searchers** who are either:

a) unfamiliar with the domain of their goal (i.e. need to learn about the topic in order to understand how to achieve their goal)

b) unsure about the ways to achieve their goals (either the technology or the process)

c) or even unsure about their goals in the first place.

Consequently, exploratory search covers **a broader class of activities** than typical information retrieval, such as ***investigating***, ***evaluating***, ***comparing***, *and* ***synthesizing***, where new information is sought in a defined conceptual area; exploratory data analysis is another example of an information exploration activity. Typically, therefore, such users generally **combine querying and browsing** strategies to foster learning and investigation."

THEREFORE…

- Ranking is not enough for exploratory search

By Gary Marchionini

EXPLORATORY SEARCH: FROM FINDING TO UNDERSTANDING

Research tools critical for exploratory search success involve the creation of new interfaces that move the process beyond predictable fact retrieval.

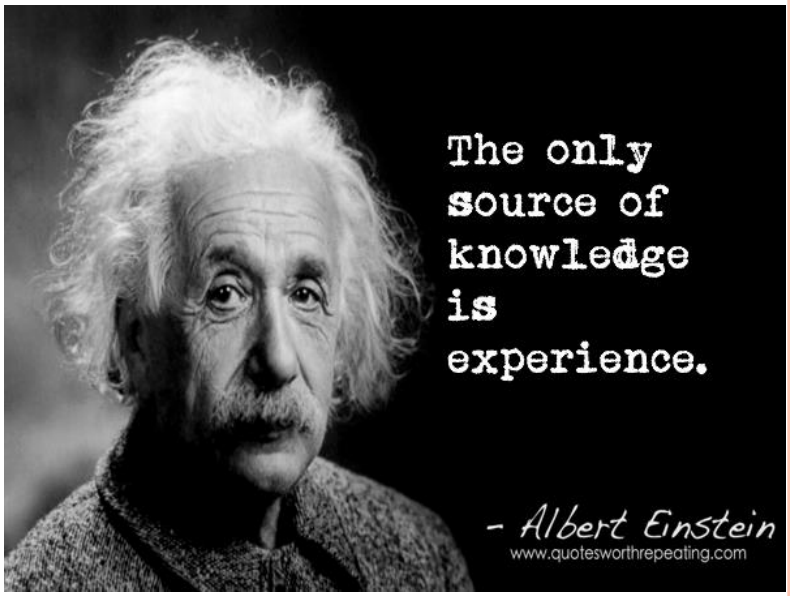
From the earliest days of computers, search has been a fundamental application that has driven research and development. For example, a paper published in the inaugural year of the *IBM Journal* 36 years ago outlined challenges of text retrieval that continue to the present [4]. Today's data storage and retrieval applications range from database systems that manage the bulk of the world's structured data to Web search engines that provide access to petabytes of text and multimedia data. As computers have become consumer products and the Internet has become a mass medium, searching the Web has become a daily activity for everyone from children to research scientists.

Some Common Requirements for Effective Exploratory Search

- Allow easy and fast access even to **low ranked** hits
- Allow browsing and inspecting the found hits in **groups** (according to various criteria)
- Offer **overviews** of the search results
  - Compute and show descriptions and **count** information for the various groups, or other **aggregated** values
- Allow **gradual** restriction/ranking of the search results

Research prototypes from ISL related to Exploratory Search

- They are presented like a “story” organized in milestones that correspond to activities of ISL (Information Systems Laboratory) of FORTH-ICS (2009-now)

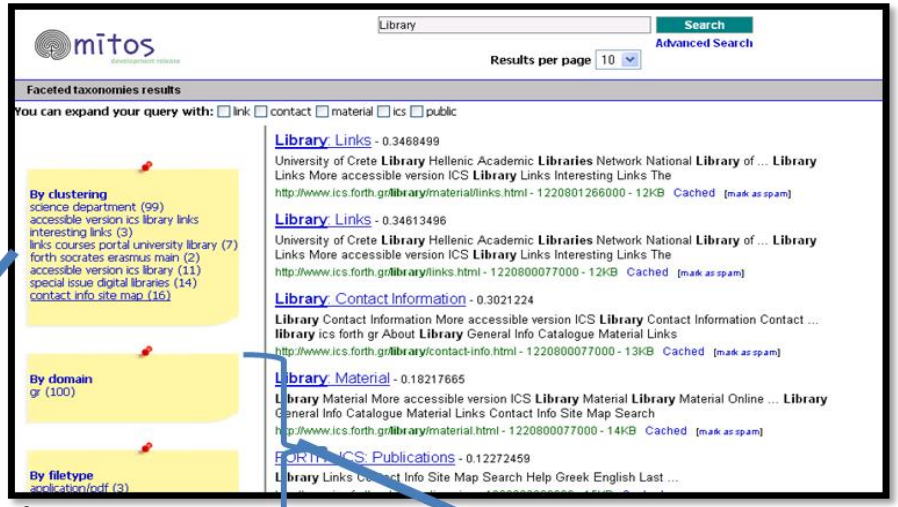


MILESTONE 1. THE MITOS WSE (2009)

mitos
development release

- MITOS is WSE built from scratch. Apart from the classical WSE functionality, Mitos offers faceted search over the results of the submitted queries.
  - It supports facets corresponding to metadata attributes of the web pages (static metadata), as well as facets corresponding to the outcome of snippet-based clustering algorithms (a kind of dynamic metadata).
  - The user can then restrict his/her focus gradually, by interacting with the resulting multidimensional structure through simple clicks.

THE MITOS WSE (2009)



Dimension based on dynamic (query-dependent) metadata (of the top ranked hits)

By clustering
- architecture (6)
- contact (10)
- content (10)
- copyright notice csd (13)
- course content english (7)
- csd (14)
- department (8)
- forth (38)
- health telematics network (6)
- ics (39)
- information (11)
- network (8)
- physical (5)
- science (22)
- ITE τεχνικές αναφορές (22)

REST (3981)

Dimensions based on static metadata

By domain
- gr (4067)

By date
- 2008 (479)
- 2007 (694)
- 2006 (1340)
- 2005 (184)
- 2004 (106)
- 2003 (82)
- 2002 (88)
- 2001 (28)
- 2000 (13)
- 1999 (4)
- 1998 (6)
- 1997 (1)
- Unknown (1042)

By filetype
- application/msword (16)
- application/pdf (1476)
- application/vnd.ms-powerpoint (29)
- text/html (2546)

By language
- Any (UTF-8) (18)
- Greek (1209)
- Latin-1 (Europe, Latin America, Caribbean, Canada, Africa) (944)
- Latin-2 (Central and Eastern Europe) (4)
- Unknown (1892)

N. Manolis and Y. Tzitzikas (ESWC'11)

10

THE MITOS WSE (2009)

A user wants to get information about **Information Systems Laboratory**

mitos development release

information systems laboratory | Search

Advanced Search

Results per page 10

Faceted taxonomies with on-demand clustering results

8558 Results 1 - 10 from 8558 for information ... ms)

8558 initial results

By clustering
- activities (39)
- biomedical informatics laboratory (10)
- decision support systems (2)
- dimitris (7)
- events (43)
- forth (25)
- history (35)
- ics (85)
- informatics (64)
- laboratories (2)
- projects (53)
- publications (47)
- seminars (26)
- support (5)
- yannis (3)

REST (8458)

By domain
- gr (8558)

By date
- 2009 (606)
- 2008 (668)
- 2007 (3113)

Information Systems Laboratory: - 0.916549

information systems laboratory ics isl img isl src images buttons isl ...

http://www.ics.forth.gr/isl/publications/by_name.jsp?Person_ID=7 - 0 - 8KB Cached - Similar pages [mark as spam]

FORTH - ICS: **Information Systems Laboratory** - 0.8703994

information systems information retrieval **systems** database workflow management **systems** semantically rich ... forth ics **information systems laboratory information systems laboratory** head **laboratory** prof

http://www.ics.forth.gr/isl/index.html - 1173087253000 - 17KB Cached - Similar pages [mark as spam]

Information Systems Laboratory: - 0.8702642

information systems laboratory ics isl panos constantopoulos muse multimedia accessible version ... ics isl isl centre cultural informatics history events activities projects publications

http://www.ics.forth.gr/isl/publications/by_year.jsp?Year_of_publication=1987 - 0 - 17KB Cached - Similar pages [mark as spam]

Information Systems Laboratory: - 0.8692224

information systems laboratory null ics isl accessible version ics isl isl ... centre cultural informatics history events activities projects publications seminars people links

publication=null - 0 -

topoulos office document retrieval multos ... accessible version ics isl isl centre cultural informatics history events activities

http://www.ics.forth.gr/isl/publications/by_year.jsp?Year_of_publication=1986 - 0 -

We can focus on "By date" facet, clicking the "2009" label.

(CONT.)

mītos development release

information systems laboratory | Search | Advanced Search

Results per page 10

Results in RDF/XML

Faceted taxonomies with on-demand clustering results

606 Results 1 - 10 from 606 for information systems laboratory. (15729 ms)

By clustering
- athanasios mouchtaris (3)
- communication (15)
- distributed (13)
- dynamic (13)
- forth (11)
- home page (3)
- networks (13)
- news (5)
- oikonomou (2)
- page (5)
- presentation (3)
- publications (6)
- spring (3)
- tziritas (2)
- πε (3)

REST (558)

By domain
- gr (606)

By date
- 2009 (606)
  - June (71)
  - May (80)
  - April (212)

Information Systems Laboratory: Seminars - 0.28782406

challenge succeed transition traditional **information systems information** retrieval **systems** database workflow ... management **systems** semantically rich large scale adaptive **information systems systems** characterized

http://www.ics.forth.gr/isl/services.html - 1244639664000 - 21KB Cached - Similar pages [mark as spam]

We can further limit the results, by selecting one of the clusters (they were recomputed for the new focus)

The results of the selected group are loaded in the results' panel and all facets are updated.

CS-463 **Information** Retrieval **Systems** - 0.18768528

CS 463 **Information** Retrieval **Systems** CS 463 **Information** Retrieval **Systems** ... Course **Information** Teaching Material Lectures and Program Exercises and Assignments Grades

http://www.csd.uoc.gr/~hy463/2007/en/announcements.html - 1241012778000 - 2KB Cached - Similar pages [mark as spam]

CS-463 **Information** Retrieval **Systems** - 0.18636355

CS 463 **Information** Retrieval **Systems** CS 463 **Information** Retrieval **Systems** Spring ... **Information** Teaching Material Lectures and Program

(CONT.)

mītos development release

information systems laboratory | Search | Advanced Search

Results per page 10

Results in RDF/XML

Faceted taxonomies with on-demand clustering results

5 Results 1 - 5 from 5 for information systems laboratory. (86 ms)

By clustering
- news (5)

By domain
- gr (5)

By date
- 2009 (5)
  - June (2)
  - May (2)
  - January (1)

By filetype
- text/html (5)

By language
- Latin-1 (Europe, Latin America, Caribbean, Canada, Africa) (5)

FORTH - ICS: News - 0.053278793

information greek **information** greek **information** greek **information** greek **information** greek **information** greek **information** ... greek **information** greek **information** greek **information** greek **information** greek **information** greek

http://www.ics.forth.gr/news/news-prev.html - 1241420849000 - 54KB Cached - Similar pages [mark as spam]

FORTH - ICS: News - 0.040669773

laboratories publications services library links contact info site map search help ...

http://www.ics.forth.gr/news.html - 1244102246000 - 23KB Cached - Similar pages [mark as spam]

FORTH - ICS: Welcome Note by the Director of ICS-FORTH - 0.028257346

zoomin ics announcements news press releases **laboratories** publications services library links ...

http://www.ics.forth.gr - 1232022089000 - 19KB Cached - Similar pages [mark as spam]

FORTH - ICS: News - 0.02504004

technical aspects multimodal **systems** tams department informatics university hamburg germany university ... **information** science university pennsylvania professor head cognitive informatics **laboratory** laval university

http://www.ics.forth.gr/news/lectures-prev.html - 1244102254000 - 100KB Cached - Similar pages [mark as spam]

FORTH - ICS: Lectures - 0.01583067

announcements news **laboratories** publications services library links contact info site map ...

http://www.ics.forth.gr/news/jan_cernocky_lecture.html - 1241420696000 - 20KB

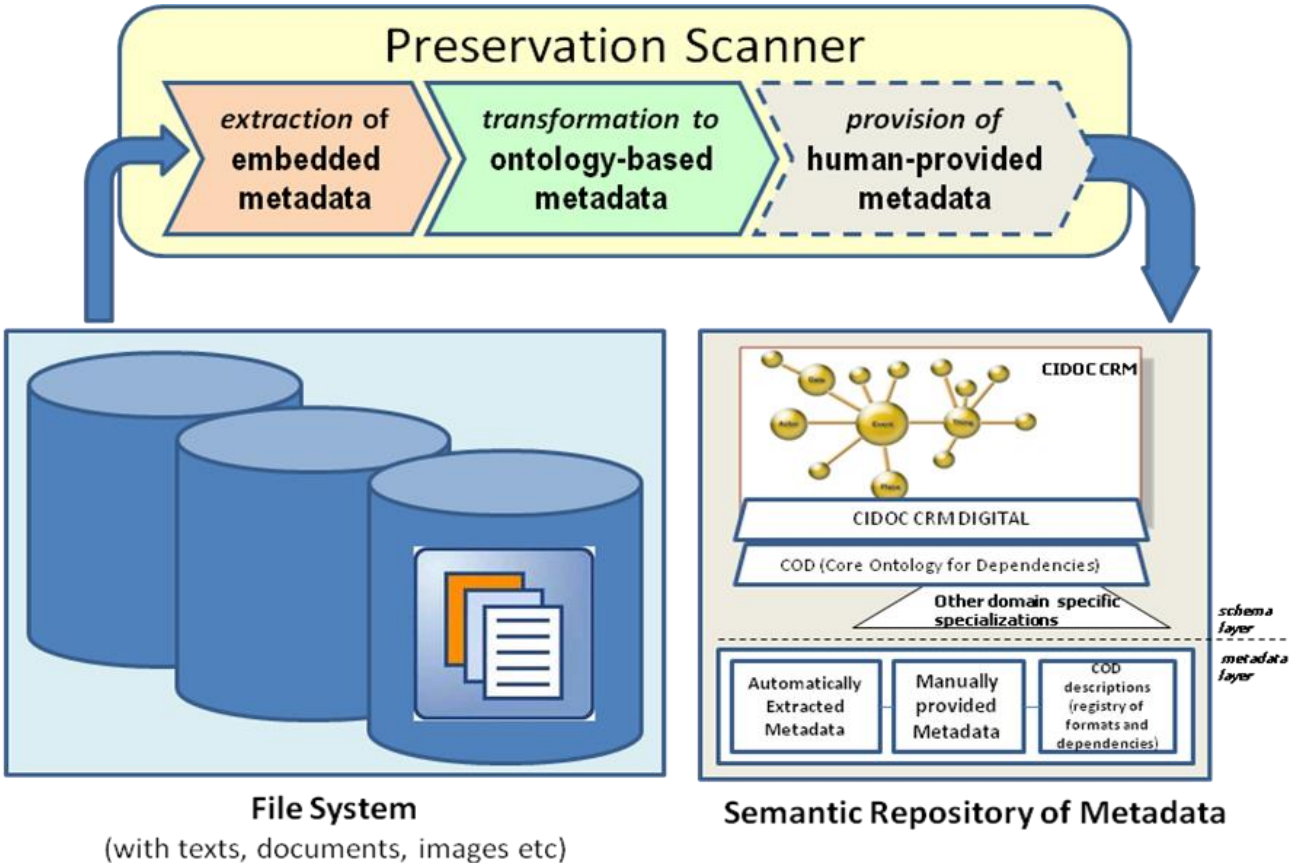
With only 2 clicks, we have limited the results to 5 hits.

CONT.

- **Evaluation** with Users (main results) :
  - Faceted search, combining dynamically and statically mined metadata
    - lead to much **improved task completeness** with much **less user interactions**
    - was **more preferred** by the users (advanced and plain ones) and lead **to greater satisfaction**, than plain clustering or faceted interfaces
- Most Important Related Publications
  - [ECDL'09] P. Papadakos, S. Kopidaki, N. Armenatzoglou and Y. Tzitzikas. Exploratory Web Searching with Dynamic Taxonomies and Results Clustering. In ECDL 2009
  - [WISE'09] S. Kopidaki, P. Papadakos, and Y. Tzitzikas. STC+ and NM-STC: Two novel online results clustering methods for web searching. In WISE 2009
  - [J. KAIS 2012] P.Papadakos, S.Kopidaki, Nikos Armenatzoglou and Y. Tzitzikas On exploiting Static and Dynamically mined Metadata for Exploratory Web Searching , KAIS Journal, 2012

PreScan

- **PreScan**: Automated extraction of file-embedded metadata from file systems.



More about PreScan

- Features
  - Automatic Scanning of file systems
  - Automatic Format Identification and Extraction of Embedded Metadata
  - Support for Human-entered/edited Metadata
  - Periodic Re-Scannings without loosing the human-provided metadata
  - Referential Integrity services
- An early version of this tool is described in the paper http://users.ics.forth.gr/~tzitzik/publications/Tzitzikas_2009_MEDES.pdf

MILESTONE 2. DURING TYPING?

<iOS>
instant overview search

- Then we questioned ourselves:
  - ***why not offering this functionality during query typing, i.e. a kind of richer autocompletion service?***
- This resulted to what we called **Instant Overview Search (IOS)**.
- The idea:
  - For the frequent queries, pre-compute and store not only the first page of results, but also the analysis of these hits
- Technical challenge
  - Since the amount of information that has to be stored for each query is higher (and obviously does not fit in main memory) we devised a partitioned trie-like index for efficiency (plus a dedicated cache)

IOS (INSTANCE OVERVIEW SEARCH), 2011-2012

This is Pavlos Fafalios
(I supervise his PhD)

We want to find information about the life of Marilyn Monroe
(and probably its connection to Pavlos?)
However, we are not sure for the spelling of her name.
So, we start typing "`mari`".

(CONT.)

home page • visit uoc-csd • visit ics-forth

» clusters overview:

- marilyn(99)
  - monroe(33)
    - quotes(4)
    - biography(3)
  - manson(8)
  - home(8)
  - free(7)
    - encyclopedia(3)
  - quotes(5)
  - photo(5)
  - collection(5)
  - news(5)
  - encyclopedia(4)
  - biography(4)
  - new(10)
    - york(4)
  - gavin rossdale(3)
  - images(5)
  - andy(3)
  - online(3)

<iOS>

instant overview search

mari | search

marilyn
marilyn monroe
mario games

ch engine
» over linked open data

- SET (default)
- PET
- STIE
- PTIE

List of query's suggestions.

First page of results of the top suggestion "marilyn"

» first page results overview:

Marilyn (singer) - Wikipedia, the free encyclopedia - 0

Peter Robinson (born 3 November 1962), better known as **Marilyn**, is a British pop singer who achieved international fa me in the 1980s with his hit song ...Early life - Blitz years - Career - Recent activity

en.wikipedia.org/wiki/Marilyn_(singer)

Marilyn Monroe - Wikipedia, the free encyclopedia - 1

Marilyn Monroe born Norma Jeane Mortenson, but baptized Norma Jea Miller - Somethings Got to Give - Some Like It Hot

en.wikipedia.org/wiki/Marilyn_Monroe

Marilyn (hill) - Wikipedia, the free encyclopedia - 2

A **Marilyn** is a mountain or hill in the United Kingdom, Ireland or Isle of ...

en.wikipedia.org/wiki/Marilyn_(hill)

Agency - 3

model agency in Paris. **Marilyn** Agency est la plus importante agence de mannequin Γ

We can continue typing the query. Instantly new suggestions are shown

Cluster Label Tree of the top suggestion "marilyn"

(CONT.)

home page • visit uoc-csd • visit ics-forth

» clusters overview:

- marilyn monroe(100)
  - quotes(8)
  - news(10)
  - online(5)
  - photos(7)
  - collection(6)
  - images(7)
  - death(6)
  - gallery(5)
  - photo(6)
  - video(4)
  - free(3)
  - pictures(5)
  - encyclopedia(3)
  - biography(5)
  - [illegible]8)

<iOS>

instant overview search

marilyn monroe | search

We selected the suggestion "`marilyn monroe`". The results' first page and cluster label tree for this suggestion were loaded immediately.

» first page results overview:

Marilyn Monroe Biography from Who2.com

Marilyn Monroe s sex appeal talent and untimely death combined to make her an enduring star and one of Hollywood s most recognizable icons. Early in.

http://www.who2.com/marilynmonroe.html

Marilyn Monroe: Biography from Answers.com

Marilyn Monroe Actor Born: 1 June 1926 Birthplace: Los Angeles California Died: 4 August 1962 (drug overdose) Best Known As: Hollywood s most.

http://www.answers.com/topic/marilyn-monroe

The Marilyn Pages-Marilyn Monroe biography and images

[illegible]life of Marilyn Monroe Norma Jean. ... Marilyn Monroe. The Marilyn Pages have moved to ellensplace.net/marilyn.ht [illegible] (If you are not taken to the new ...

[illegible]//www.ionet.net/~jellenc/marilyn.html

[illegible]e Marilyn Pages-Marilyn Monroe biography and images

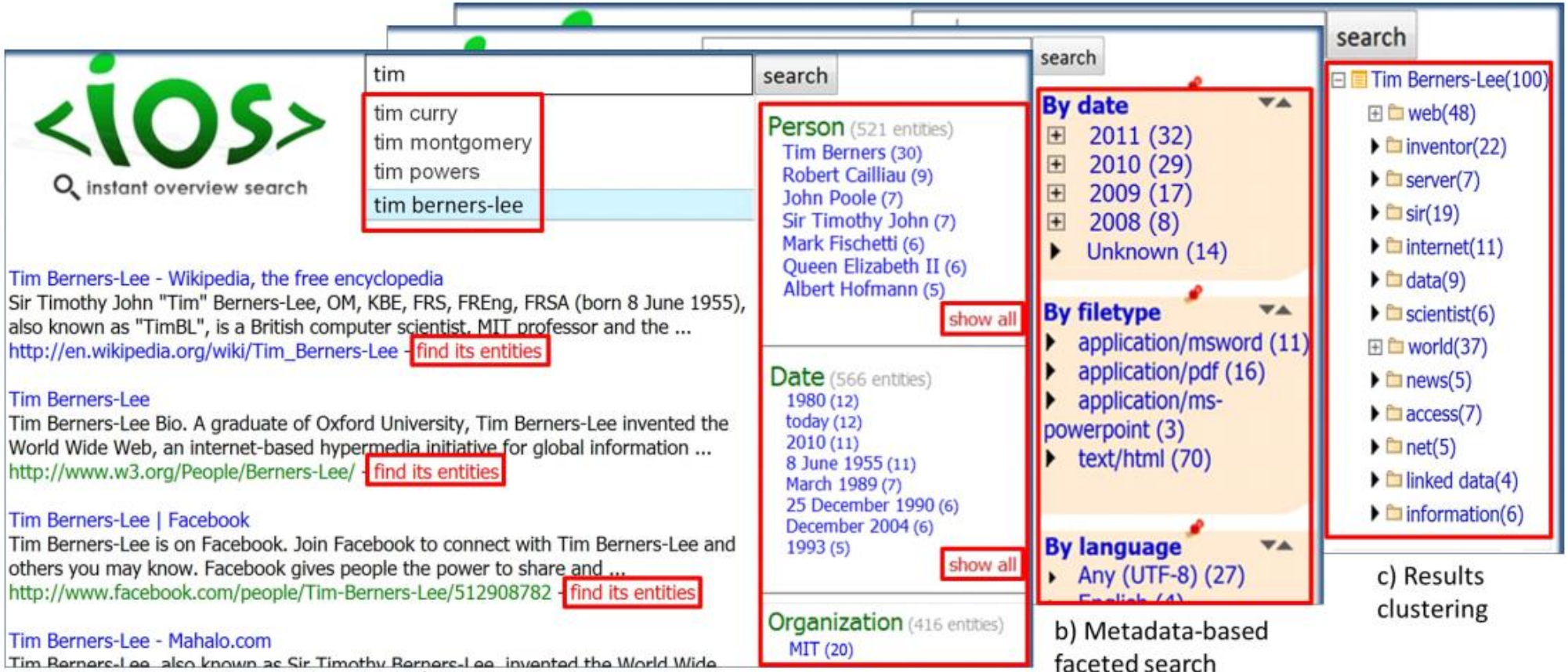
[illegible]life of Marilyn Monroe Norma Jean. ... NOTE FOR AOL USERS ◆ Site Awards for The Marilyn Pages ◆ to ellen s [illegible]e.

http://www.ellensplace.net/marilyn.html

By clicking a label, the results of the specific cluster are loaded in the results panel.

IOS (INSTANCE OVERVIEW SEARCH)

We can exploit this technique <u>for any king of pre-processing of search results</u> (e.g. metadata-based faceted search, snippet-based clustering, entity mining, etc)

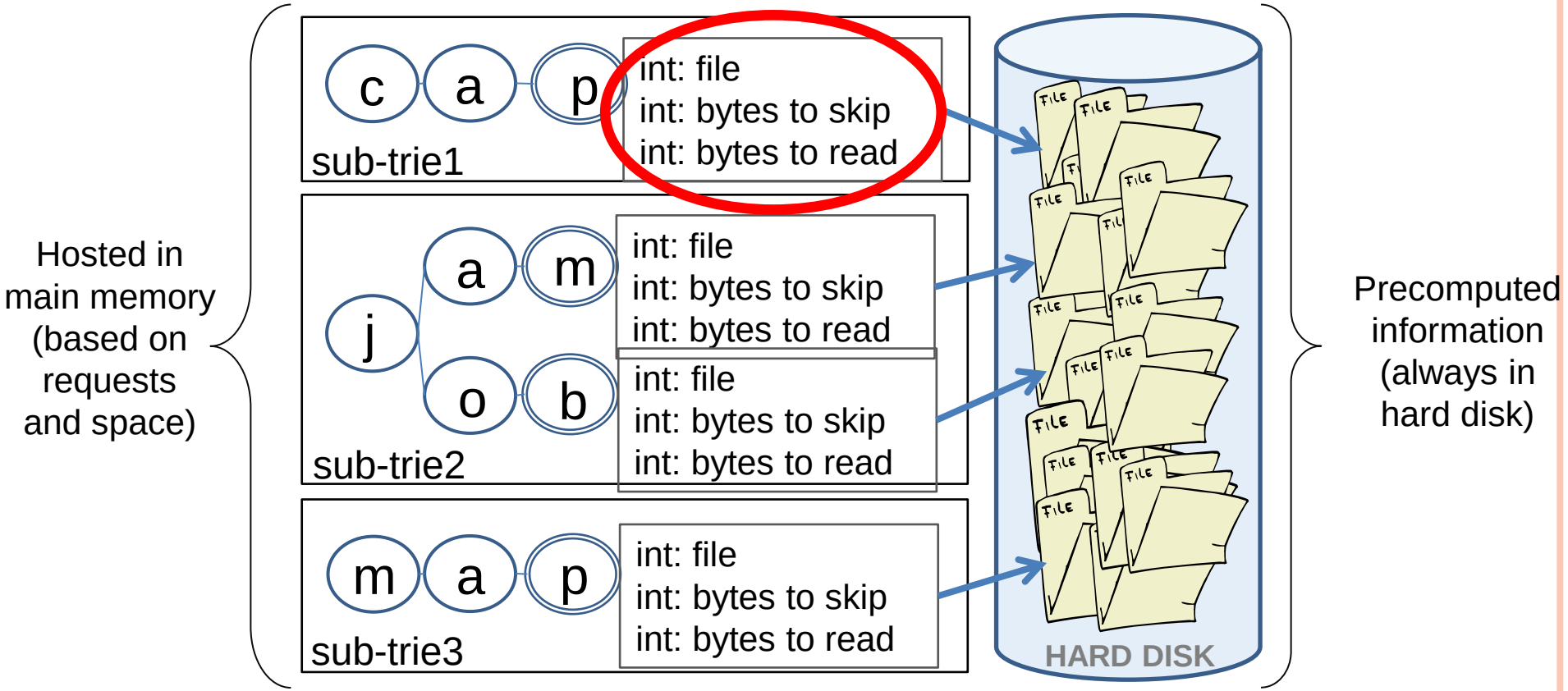


a) Entity mining

b) Metadata-based faceted search

c) Results clustering

IOS Indexes



Average Retrieval Time ≈ **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***

CONT

- Key results
  - A partitioned trie-based index structure that can efficiently support recommendations for millions of distinct queries even with modest hardware
    - One can provide instant access to large amount of data, utilizing the existing resources, without requiring more hardware
  - A hybrid caching policy (70% static and 30% dynamic) seems to be the more appropriate choice yielding a throughput increment of around 80% and a 25% speedup
- Demo
  - http://www.ics.forth.gr/isl/ios
    - Select the system "Instant Entity Mining + Clustering (over Bing)"
- Related Publications

- [WISE'11] P. Fafalios and Y. Tzitzikas, Exploiting Available Memory and Disk for Scalable Instant Overview Search, 12th International Conference on Web Information System Engineering (WISE 2011), Sydney, Australia, October 2011
- [WWW'12] P. Fafalios, I. Kitsos and Y. Tzitzikas, Scalable, Flexible and Generic Instant Overview Search, 21st International Conference on World Wide Web, (WWW 2012), Demo Paper, Lyon, France, April 2012

MILESTONE 3. ENTITY MINING AND LOD?

- Then we questioned ourselves:
  - ***why not exploiting LOD in the context of entity mining of the search results?***
- Motivation
  - LOD contains plenty of information about Named Entities (their names, attributes, relationships with other entities, etc)
- Output
  - IOS Entity Mining
    - LOD is used as source for Named Entity Recognition
    - LOD is used for providing more information about the identified entities

IOS Entity Mining (2012)

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

entity mining IOS

barack obama

Search

100 results to mine

mine only snippets

Barack Obama - Wikipedia, the free encyclopedia
Barack Hussein Obama II (born August 4, 1961) is the 44th and current President of the United States. He is the first African American to hold the office. Obama ...
http://en.wikipedia.org/wiki/Barack_Obama - find its entities

Barack Obama
Barackobama.com is the official re-election campaign website of President Barack Obama. Visit the site for the latest updates from

Quote of the day: "This is a good first step, but it is only a step. Congress needs to pass the rest of my American Jobs Act so that can create jobs and put ...
https://www.facebook.com/barackobama - find its entities

Person (1427 entities)
Barack
Michelle
George
Ann Du
Craig R
Joe Bid
John M
Kennedy
Sarkozy (8)
Clinton (6)
show all

Nicolas Sarkozy
Current President of France
Birth date: 1955-01-28
Birth place: Paris, France
Profession: Lawyer
Web site: http://www.sarkozy.fr
Page: http://en.wikipedia.org/wiki/Nicolas_Sarkozy

Organization (842 entities)
Harvard (14)
White House (18)
Congress (14)
University of Hawaii (10)
Columbia University (8)

- Automatically connects knowledge with documents at query time
- No preprocessing
- No indexing

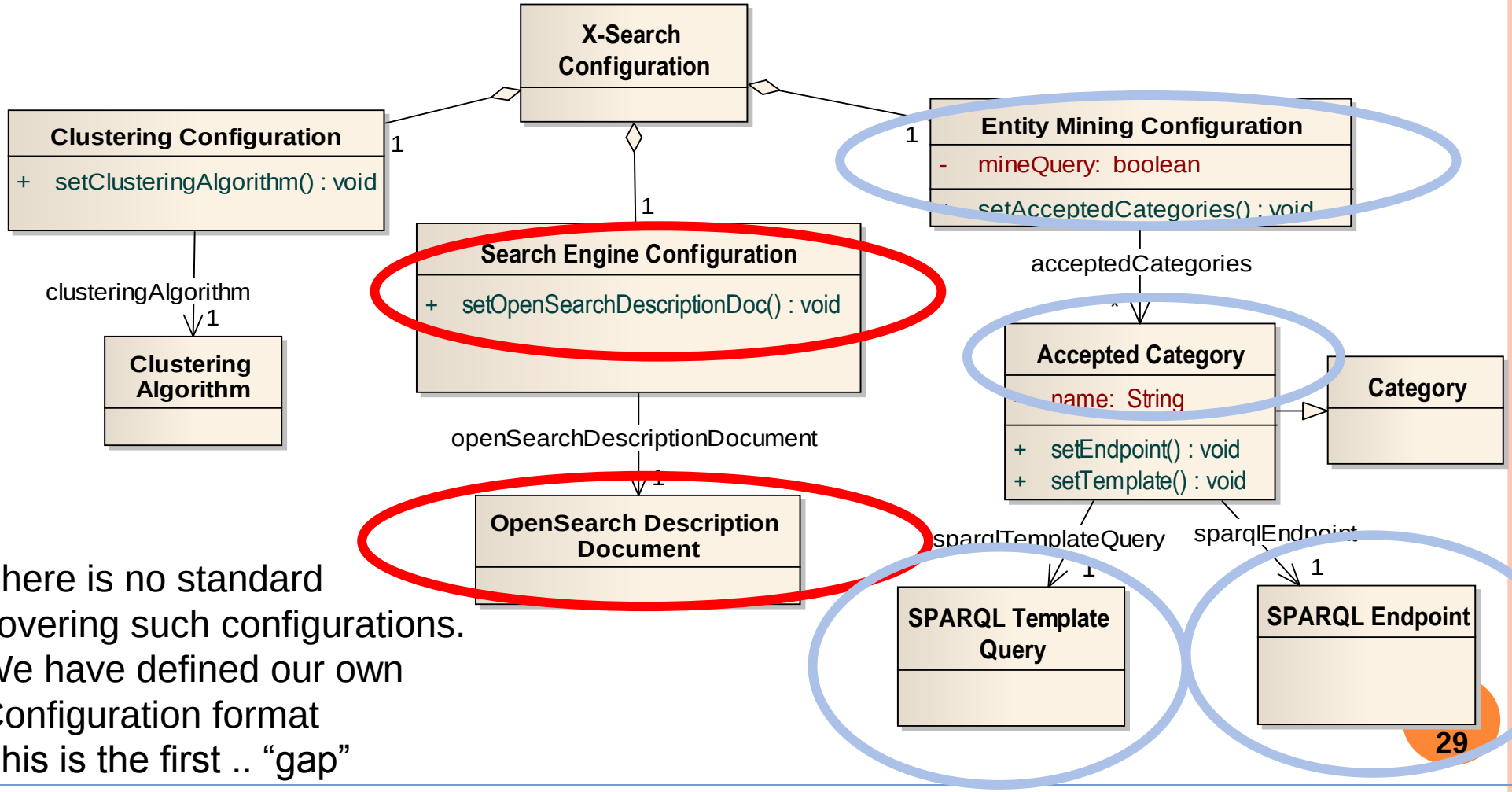
CONT

- Exploitation for restricting the focus

MILESTONE 4. CONFIGURABILITY (AND LOD)

- Then we questioned ourselves:
  - ***why not allowing the user to configure himself the entities of interest by exploiting LOD (again in the context of entity mining of the search results)?***
- Outcome
  - X-ENS (e**X**plore **EN**tities in **S**earch)
- Related Publications
  - [SIGIR'13] P. Fafalios and Y. Tzitzikas, X-ENS: Semantic Enrichment of Web Search Results at Real-Time, 36th International ACM SIGIR Conference, Demo Paper, Dublin, Ireland, 28 July - 1 August 2013

XSEARCH-CONFIGURABILITY: THE CONCEPTUAL MODEL



There is no standard covering such configurations.
We have defined our own Configuration format
This is the first .. “gap”

X-ENS (SIGIR 2013)

http://139.91.183.72/x-ens

MILESTONE 5. PROFESSIONAL SEARCH SYSTEMS?

- Then we questioned ourselves:
  - ***why not applying and testing this in the context of a professional search system?***
- Outcome
  - Application in **patent search**. Missing relevant documents is unacceptable in patent search *(recall oriented search procedure)*. Retrieval of all relevant documents is usually necessary
    - Patents contain plenty of named entities of various kinds
      - *Companies, Countries, Persons, Product types, Laws, etc*
  - Inclusion of **PerFedPat** System
    - In collaboration with Mike Salampasis

PERFEDPAT

http://www.perfedpat.eu/

ezDL - (visitor)

File Tools Perspectives Help

Advanced Query

Basic Search
Patent Search
Bibliographic Search
Classification Search
Search Options

Full Text/Abstract: migraine
Title: e.g. plastic A...
Publication number: e.g. WO2008...
Application number: e.g. DE19971...
Priority number: e.g. WO1995...
Year: e.g. >=2005 A...
Applicant(s): e.g. "Institut P...
Inventor(s): e.g. Smith
European Classification (ECLA): e.g. F03G7,10
International Patent Classification (IPC): e.g. H03M1,12
U.S. Classification: e.g. 280,251

Clear Search

Entities Explorer

Refresh

Powered by X-Search

International Patent Classification (IPC), Inventor, Applicant, European Classification (ECLA), Disease, Publication Year, Publication Country, Application Year, Application Country, Publication Number, Drug, Chemical Substance

International Patent Classification (IPC) (111 entities)
A61P25/00 (28)
A61P25/06 (23)
A61P25/04 (11)
A61K31/185 (6)
A61K31/445 (6)
C07D451/00 (6)
A61K31/46 (5)
A61K31/55 (4)
A61K31/403 (5)

Cluster Ex...

Refresh

Powered by X-Search

text:migraine(50)
- migraine(46)
- behandlung(29)
- traitement(28)
- migräne(24)
- treatment(23)
- verwendung(13)
- treating(14)
- method(12)
- prevention(8)
- verfahren(8)
- vorbeugung(8)
- pain(7)
- douleur(6)
- traiter(7)

Results, Details | Library Choice

Results

Results: 200

Relevance | Enter filter terms

Gro... | Nothing

1. Use of nadolol for inhibiting the onset of migraine
SQUIBB & SON & INC
IPCs: [A61K31/22, A61K31/21, A61K31/135] EP-0350080-A2, 1

2. Method for treating migraine symptoms with ibup
MCNEIL PPC INC

Details

USE OF NADOLOL FOR INHIBITING THE ONSET OF MIGRAINE HEADACHES

Year: 1990

Publication number: EP0350080

Kind: A3

Application number: EP89115176

Search finished

PERFEDPAT (CONT)

- The proposed functionality:
  - offers a tight integration of different search tools with the main retrieval engine,
  - connects the search results (i.e. patents) with data and knowledge,
  - can be exploited by any patent search system (i.e. it acts as a service over a ranked list of results)
  - The time that we have to pay is proportional to the number of the top results that we want to "explore" (≈ 1.5 sec / 100 results)
- Related Publications
  - P. Fafalios, M. Salampasis and Y. Tzitzikas, Exploratory Patent Search with Faceted Search and Configurable Entity Mining, 1st International Workshop on Integrating IR technologies for Professional Search, in conjunction with ECIR'13, Moscow, Russia, March 2013

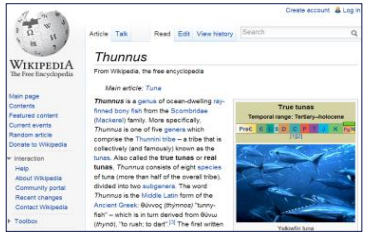
MILESTONE 6

APPLYING IN THE CONTEXT OF AN INFRASTRUCTURE

- Then we questioned ourselves:
  - ***why not applying this in another domain of professional search in the context of a real and operating EU research infrastructure?***
- Outcome
  - X-Search in the context of the ongoing iMarine Research Infrastructure project

XSEARCH:
SEMANTIC POST-PROCESSING OF SEARCH RESULTS

http://www.i-marine.eu/



web browsing

contents

query terms

(top-L) results (+ metadata)

MarineTLO Warehouse

Entity Mining

entities / contents

Semantic Analysis

- Grouping,
- Ranking
- Retrieving more properties

semantic data

Visualization/Interaction
(faceted search, entity exploration, annotation, top-k graphs, etc.)

Example: X-Search deployed in an Operational Research Infrastructure (2012-now)

Semantically Enriched Results

Query: tuna
In Collections: FIGIS

Mined Entities

- FAOCountry(24)
  - Republic of ..(1)
  - Viet nam(1)
  - Venezuela(2)
  - Yugoslavia(2)
  - Senegal(1)
  - showA
- Species(8)
  - eastern Paci..(1)
  - yellowtail a..(1)
  - Ara(1)
  - pantropical ..(1)
  - Indo-Pacific..(1)
  - showA
- WaterAreas(3)
  - Mediterranea..(1)

Object | **Metadata**

Thunnus albacares (Bonnaterre, 1788) - Fact sheet

Yellowfin **tuna**... (Venezuela),Ca bo Vang (Viet nam),**Tuna** zutoperka (Yugoslavia)...There are important yellowfin **tuna** fisheries throughout tropical and subtropical seas. The most... major surface fishing techniques for yellowfin **tuna** in the Pacific, even though this method

Textual Clustering

- Root(15)
  - fact sheet(27)
  - thunnus(8)
  - stenella(4)
  - linnaeus fac..(3)
  - xis(2)
  - tengraulis..(1)
  - dax fact s..(1)
  - purus lin..(1)

Semantic Entity Exploration

- **URI:** http://www.fao.org/figis/flod/entities/codedentity/3e6d22db-1f06-437d-ac4a-9d3c8b895bf5 (open)
- **Value:** yellowtail amberjack
- **URI:** http://dbpedia.org/re
- **Value:** Yellowtail amber

Semantic Entity Exploration

Properties of: Yellowtail_amberjack

Type	SameAs
• Animal (open)	• Seriola_lalandi (open)
• Thing (open)	
• Species (open)	
• FLODSpecies (open)	
• Fish (open)	
• CodedEntity (open)	
• Eukaryote (open)	
• Animal (open)	
• Fish (open)	

Subject
- Category:Fish_of_the_Red_Sea (open)
- Category:Fish_of_the_Indian_Ocean (open)
- Category:Seriola (open)

BinomialAuthority
- Georges_Cuvier (open)
- Achille_Valenciennes (open)

Class
- Actinopterygii (open)

Family
- Carangidae (open)

Genus
- Seriola (open)

Kingdom
- Animal (open)

Order
- Perciformes (open)

Phylum
- Chordate (open)

Depiction
- Seriola_lalandi.jpg (open)

Thumbnail
- 200px-Seriola_lalandi.jpg (open)

Object | **Metadata**

Thunnus thynnus (L

Atlantic bluefin **tuna**.. fisheries. Off Sicily, no oceanic but seasonal tolerate a... for almost catches of northern bl

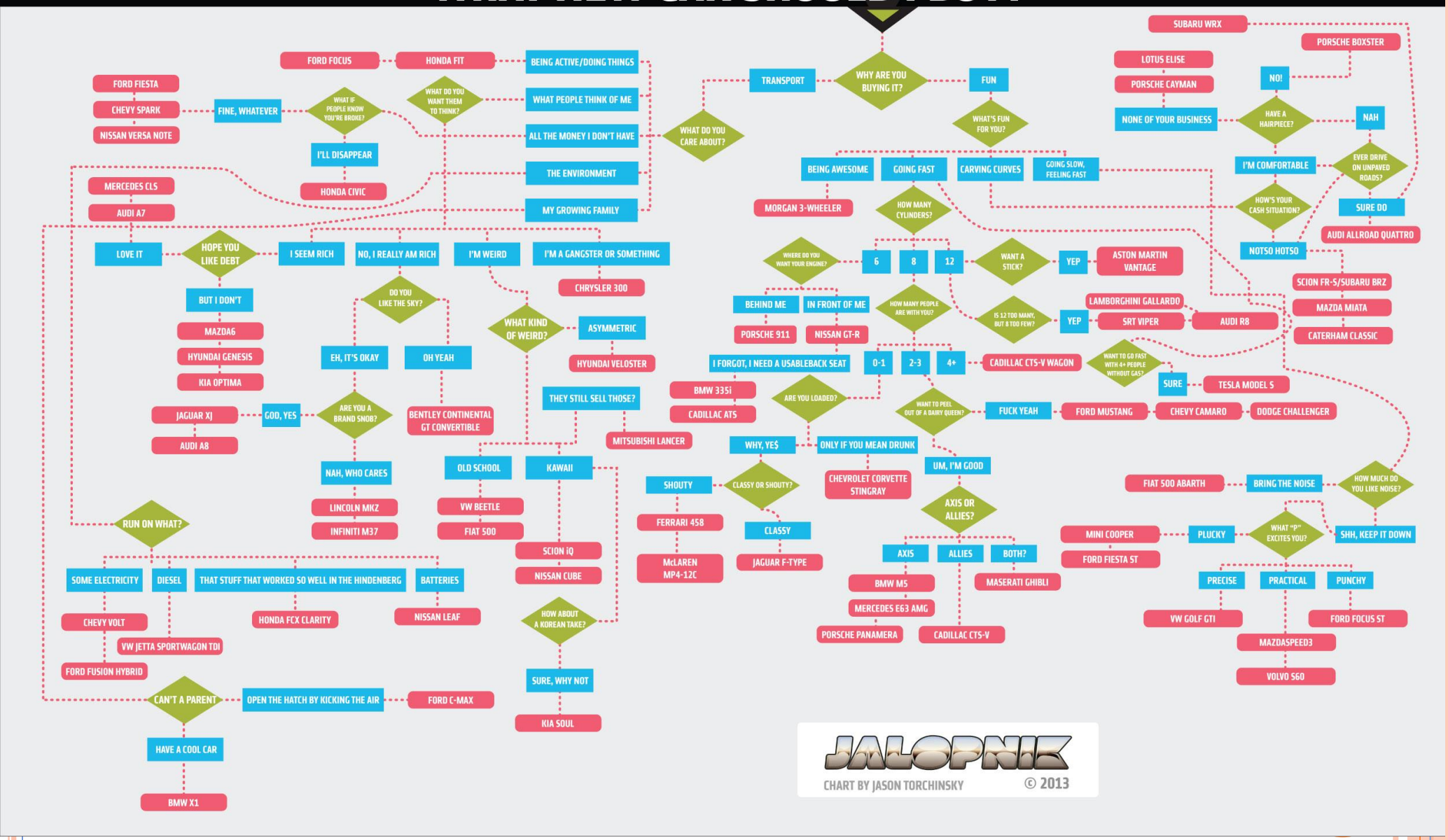
Result of Entity Mining

esult of extual stering

Y. Tzitzikas, Panel@ExploreDB 2014


RUBICON

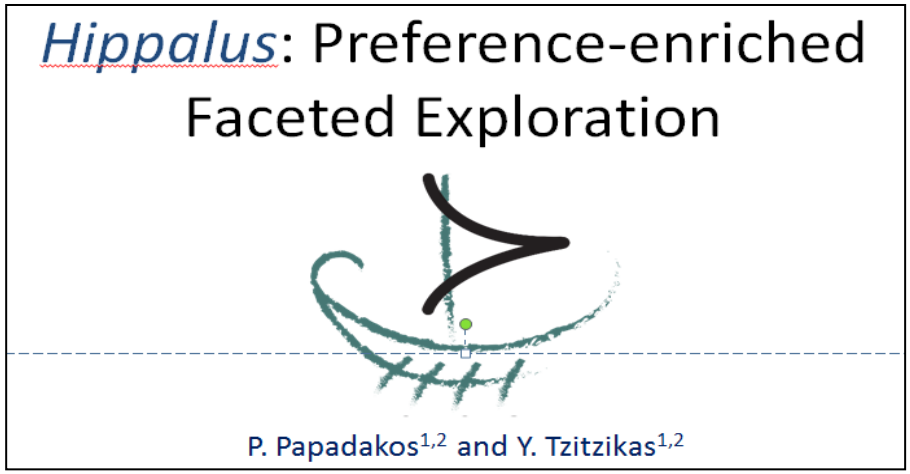
WHAT NEW CAR SHOULD I BUY?



MILESTONE 7 PREFERENCES

- Then we questioned ourselves:
  - ***What about the ordering of facets, terms and objects? Should the user only restrict the focus? Why not allowing the user to change the order based on his/her preferences?***
- Outcome
  - A framework for preferences over multi-dimensional and hierarchical information spaces
  - An extension of the interaction model of faceted search with preferences
  - The Hippalus system that realizes it

System: Hippalus (2013)



- Allows faceted browsing and also supports Preferences
  - User actions specify the ranking of the information space
  - Gradual preference specification
  - Automatic resolution of conflicts
  - Different preference composition modes
    - E.g. if the user defines the desired ordering wrt each dimension, then the first block of the ranked objects is the skyline

Hippalus: Interaction over a KB of 50 cars

Hippalus

Preference-Enriched Faceted Exploratory System

Facets

- + Acceleration (43)
- + Body_Type (50)
- + Doors (50)
- + Drive System (50)
- + Engine_Power (50)
- + Engine_Torque (48)
- + Engine_Volume (50)
- + Fuel_Cons_city (43)
- + Fuel_Cons_highway (43)
- + Fuel_Tank (46)
- + Fuel_Type (50)
- + Gears (50)
- + ID (50)
- + Manufacturer (50)
- + Model (50)
- + Price (50)
- + Speed (47)
- + Transmission (50)
- + Trunk (40)
- + Vehicle_Type (50)
- + Weight_Empty (39)
- + Year (50)

In focus: **50** objects Number of buckets: **1**

1

- Alfa-Romeo-8C-ID3
- Alfa-Romeo-Brera-ID1
- Alfa-Romeo-MiTo-ID2
- Audi-A3-ID4
- Audi-S8-ID5
- Audi-TT-ID6
- BMW-1-ID7
- BMW-3-ID8
- BMW-7-ID9
- Citroen-C1-ID10
- Citroen-C3-ID11
- Fiat-Bravo-ID12
- Fiat-Punto-ID13
- Ford-Fiesta-ID14
- Ford-Ka-ID15
- Hyundai-i10-ID16
- Hyundai-i30-ID17
- Kia-Ceed-ID18
- Lancia-Delta-ID19
- Mazda-3-ID20
- Mazda-RX-8-ID21
- Mercedes-Benz-A-ID22
- Mercedes-Benz-C-ID23
- Mercedes-Benz-C-ID25
- Mercedes-Benz-SL-ID24
- Mitsubishi-Colt-ID26
- Mitsubishi-X-Trail-ID27
- Nissan-Micra-ID28
- Nissan-Navara-ID29
- Opel-Astra-ID30

Preference Actions

Clear

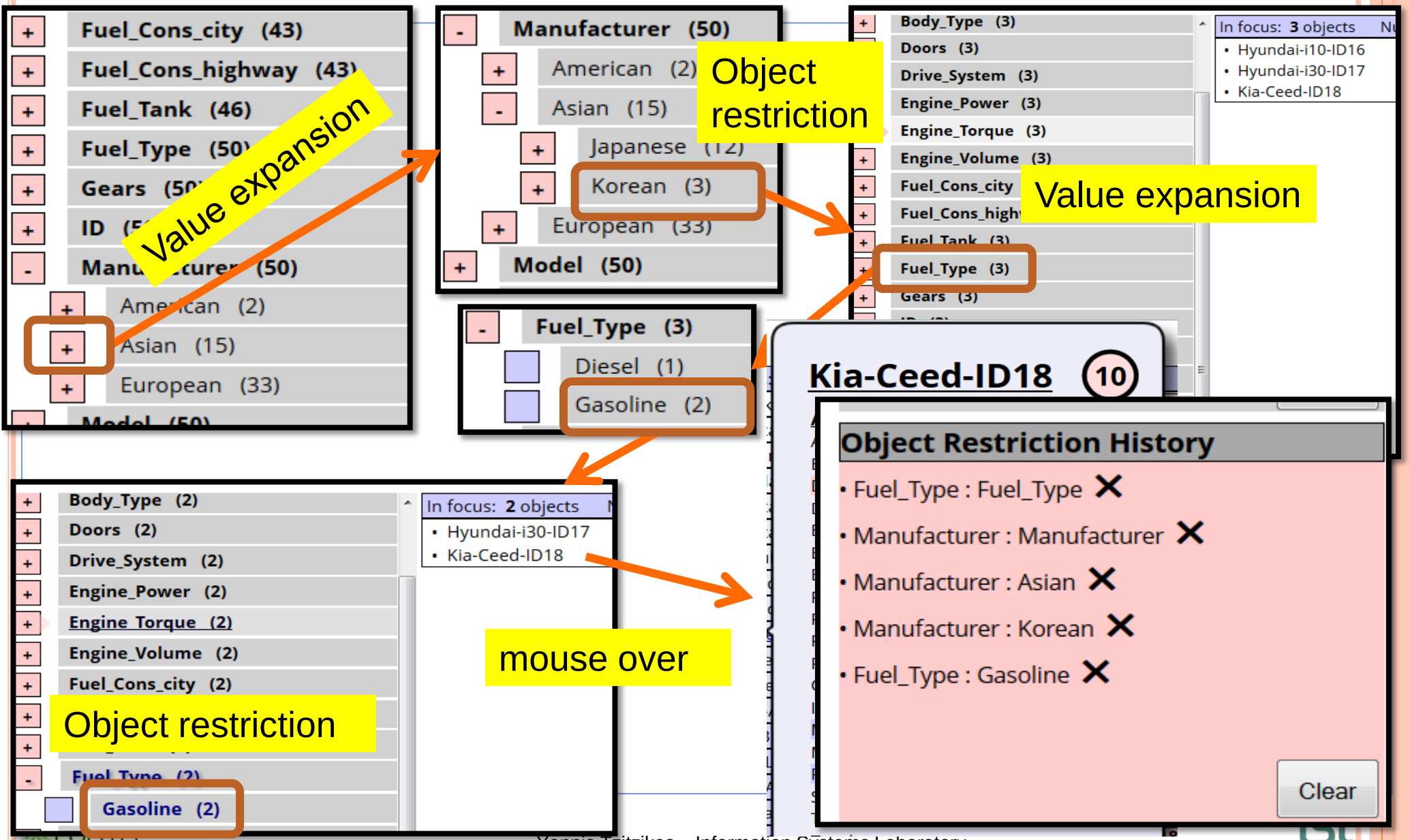
Composition: Combination

Interesting Objects

Clear

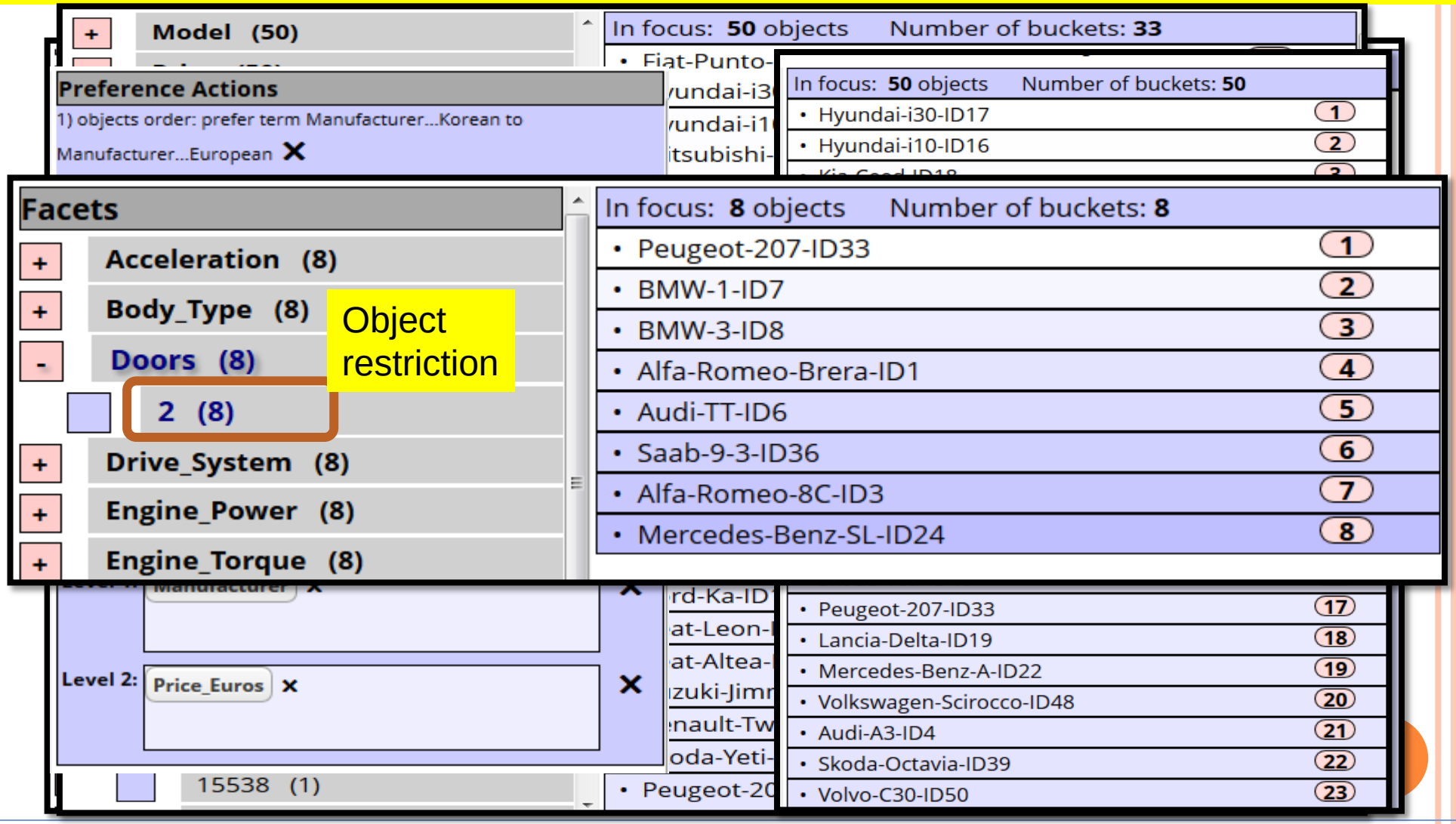
Object Restriction History

HIPPALUS: FDT INTERACTIONS



HIPPALUS: PREFERENCE ACTIONS

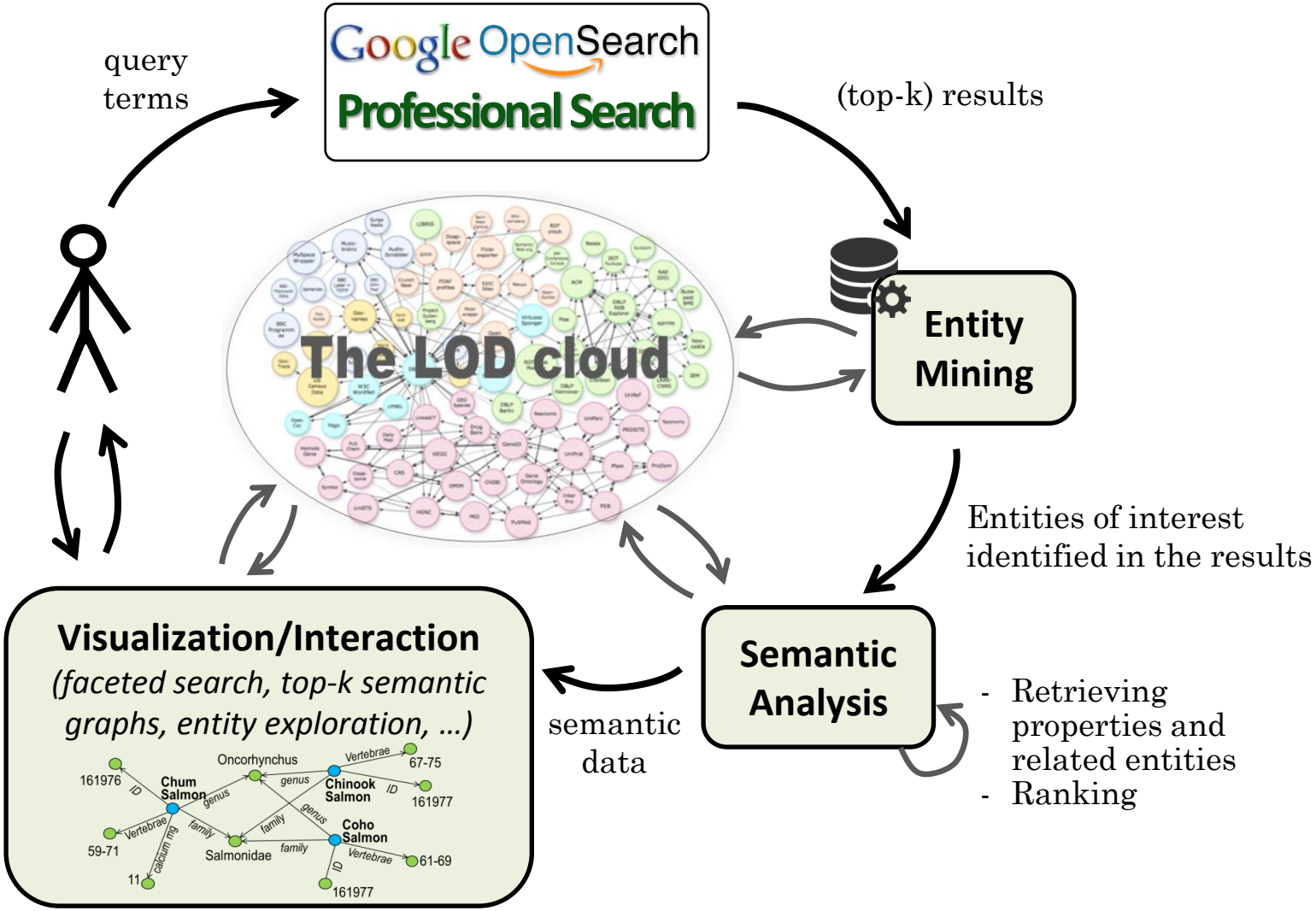
Cars ordered with priority on manufacturer



MILESTONE 8 FROM DIMENSIONS TO GRAPHS

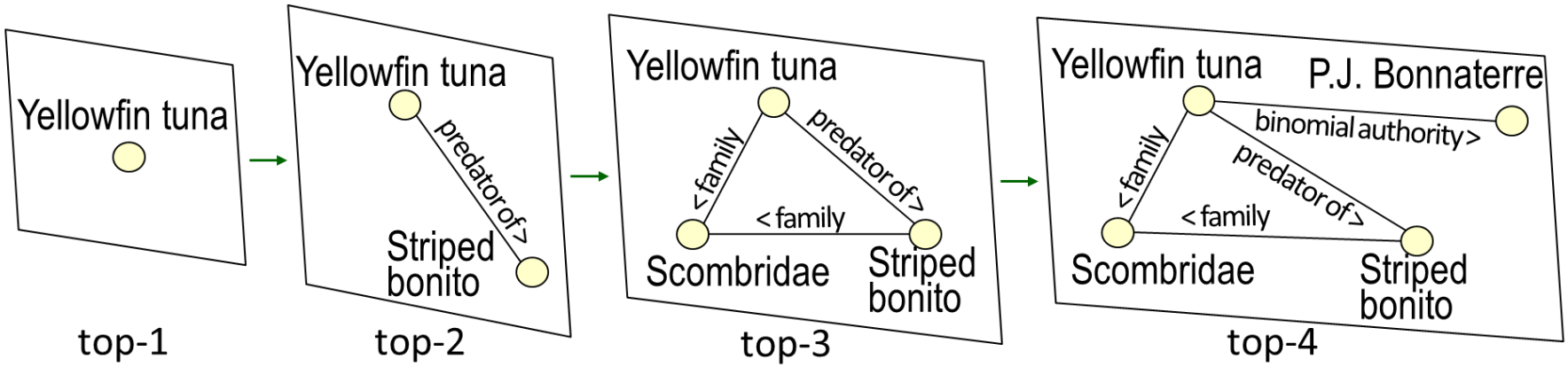
- Then we questioned ourselves:
  - ***So far we have seen services for getting and exploiting multidimensional spaces over the search results. But what if the notion of dimension cannot be defined, or in case there are too many? What can be done without having to configure entity types?***
- Outcome
  - A semantic post-processing of results that does not yield a multidimensional space but a graph.
- Challenges
  - Graph construction and exploitation for identifying the important (useful for the user) nodes and relationships

Top-K Semantic Graphs



Top-K Semantic Graphs (cont.)

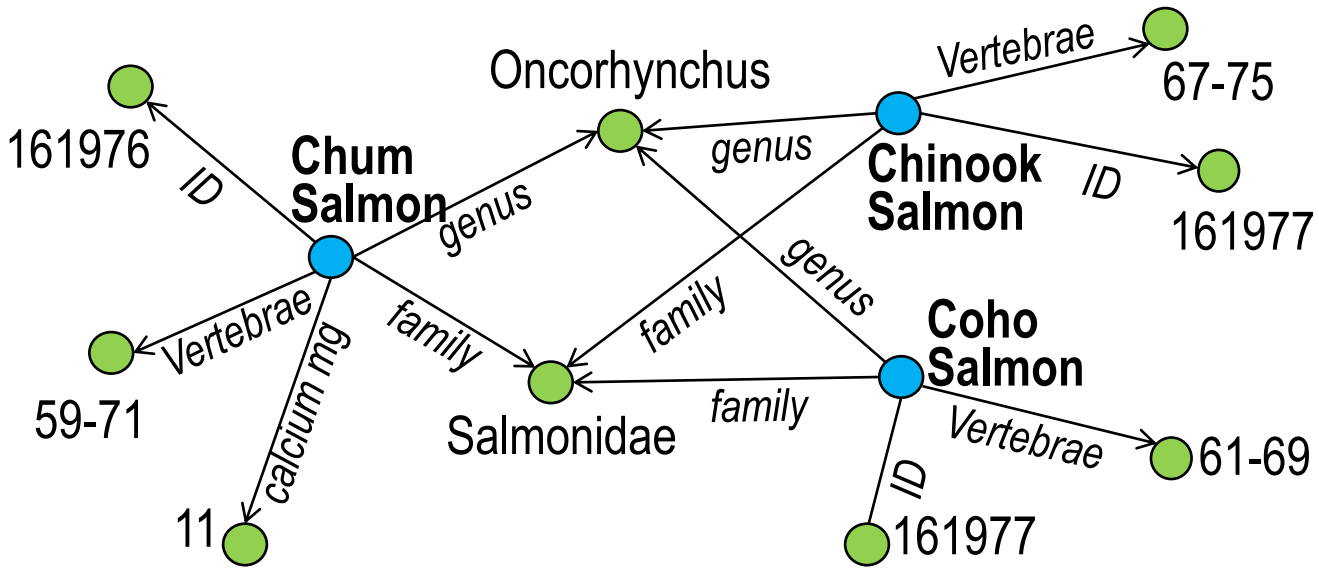
- The system can return the top-K graph for any K from 1 to number of nodes produced
  - Vertices: the K most highly ranked nodes
  - Edges: the edges that connect the K most highly ranked nodes
- The user is free to increase or reduce the value of K
- Example (from a real domain):



Top-K Semantic Graph

This graph

- can **complement** the query answer with useful information regarding the connectivity of the identified entities
- allows users to **instantly inspect** information that may lie in different places and that may be laborious and time consuming to locate
- provides useful information about the **context** of the identified entities
- allows the users to get a **more sophisticated overview** and to make better sense of the results



PROTOTYPE (2014)

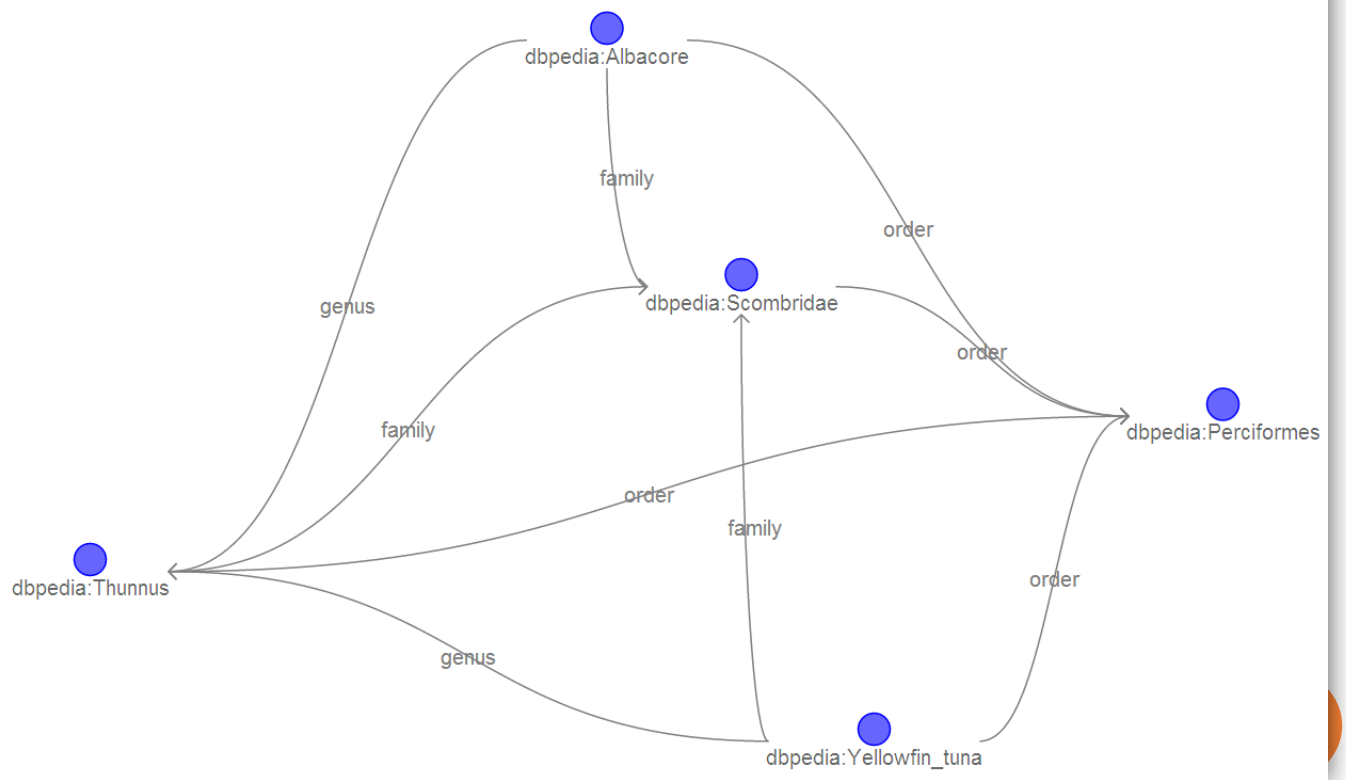
http://139.91.183.72/x-ens-2/

yellowfin tuna | Search

TOP-5 LIST

1. dbpedia:Yellowfin_tuna
2. dbpedia:Perciformes
3. dbpedia:Scombridae
4. dbpedia:Albacore
5. dbpedia:Thunnus

TOP-5 GRAPH



CONT.

- Evaluation (main results)
  - Usefulness – Survey for the marine domain
    - *The majority of participants believe that the appearance of a graph of semantic information related to the search results can help them during an exploratory search process*
  - Effectiveness – Comparative evaluation of ranking schemes:
    - The proposed PageRank-based ranking scheme produces more preferred ranking compared to other link analysis-based algorithms
  - Efficiency – Case study over online DBpedia
    - The exploitation of LOD can be supported at query-time
    - For up to 100 detected entities we can offer the proposed functionality at real-time, even if we query an online KB (like DBpedia)
  - The major bottleneck is the reliability and performance of online SPARQL endpoints
    - We expect this limitation to get overcome in the near future
    - In the meanwhile, we can use caching / indexing / dedicated warehouses / distributed infrastructure
- Related Publications:
  - P. Fafalios and Y. Tzitzikas, Post-Analysis of Keyword-based Search Results using Entity Mining, Linked Data and Link Analysis at Query Time, IEEE 8th International Conference on Semantic Computing (ICSC'14), Newport Beach, California, USA, June 2014

Hippalus Demo

- With Firefox **version 8+** try http:/www.ics.forth.gr/isl/Hippalus

CONTACT PERSON:
YANNIS TZITZIKAS
(HTTP://WWW.ICS.FORTH.GR/~TZITZIK)