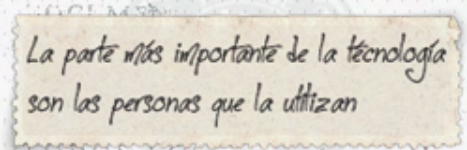
Grup de Recerca en Interacció
Persona Ordinador i Integració de Dades
fundat pel Dr. Jesús Lorés



Design and Evaluation of Overview Components for Effective Semantic Data Exploration

Josep Maria Brunetti,
josepmbrunetti@diei.udl.cat

Universitat de Lleida

Agenda

1. Introduction and Motivation
2. Approach
3. Overview Components
4. Evaluation
5. Conclusions and Future Work

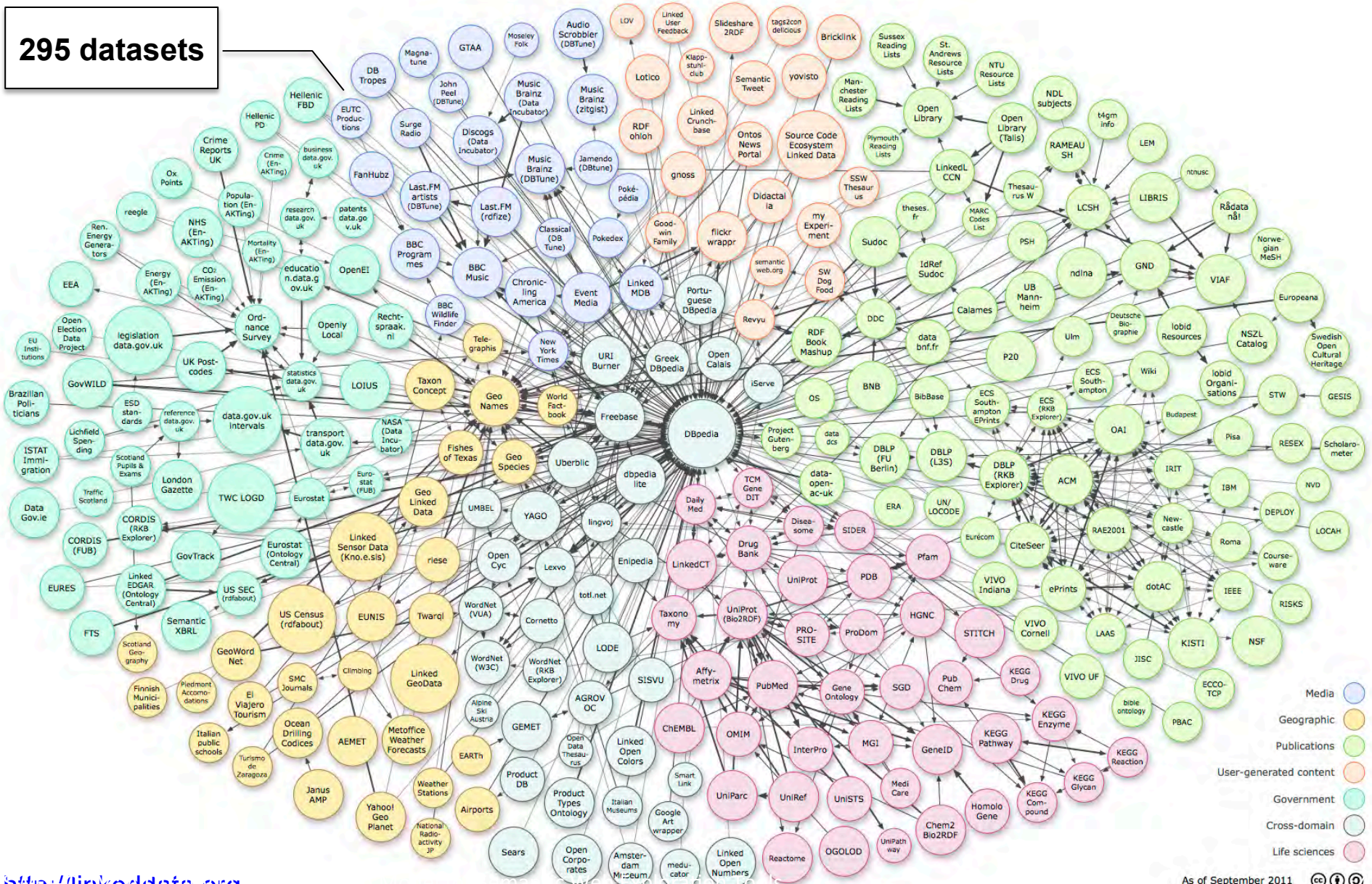
Introduction

- The Semantic Web proposed in 2001…
  …but **not popular** until recently

- **Linked / Open Data initiatives**
  - Publish Open Data in reusable formats
  - **Enormous potential**

- **But what about end users?**

Semantic Datasets



Fine for computers… but people?

C. Warren
(blogger)

I'm writing about "Films I Like".
Can I reuse LinkedMDB?

M. Harper
(developer)

I'm developing a bird watching application.
Can I reuse DBPedia?

Media
Geographic
Publications
User-generated content
Government
Cross-domain
Life sciences

As of September 2011

http://linkeddata.org

Example scenario

What to do with **DBPedia**?

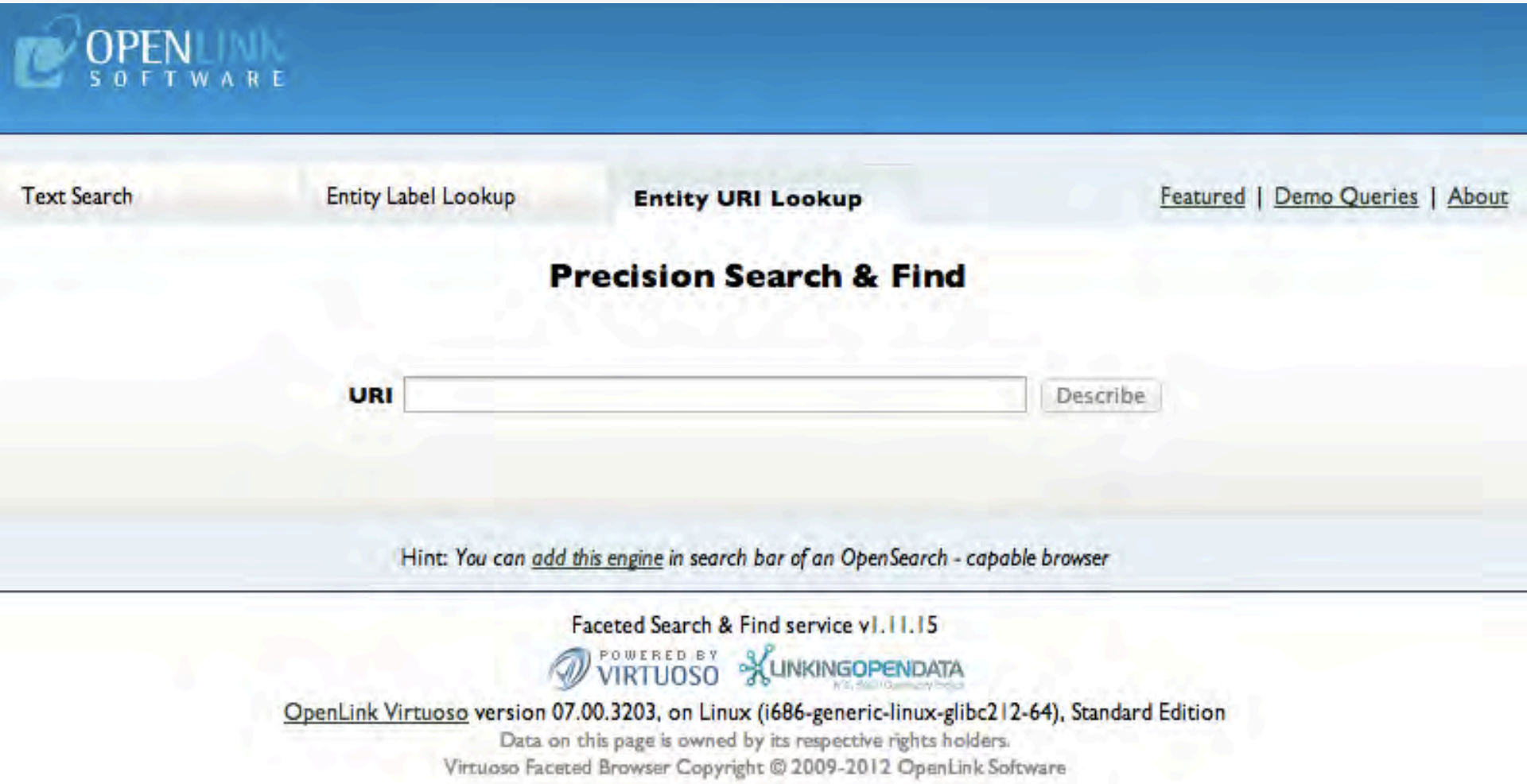
- 3.5 million things described
- Ontology: 257 classes and 1276 properties

Typical questions:

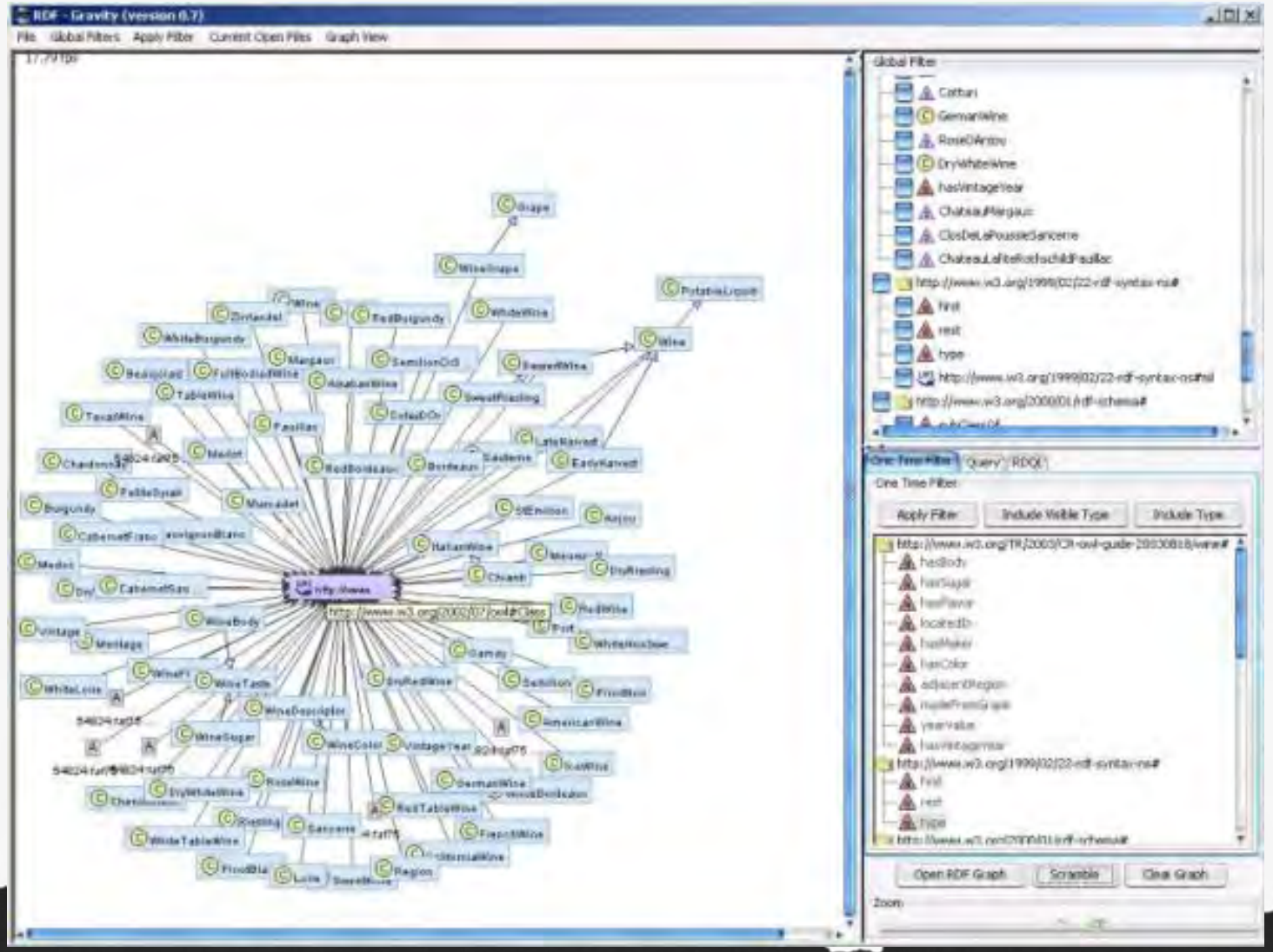
- Where do I start?
- Where do I go now?
- What is the data about?

Information Overload

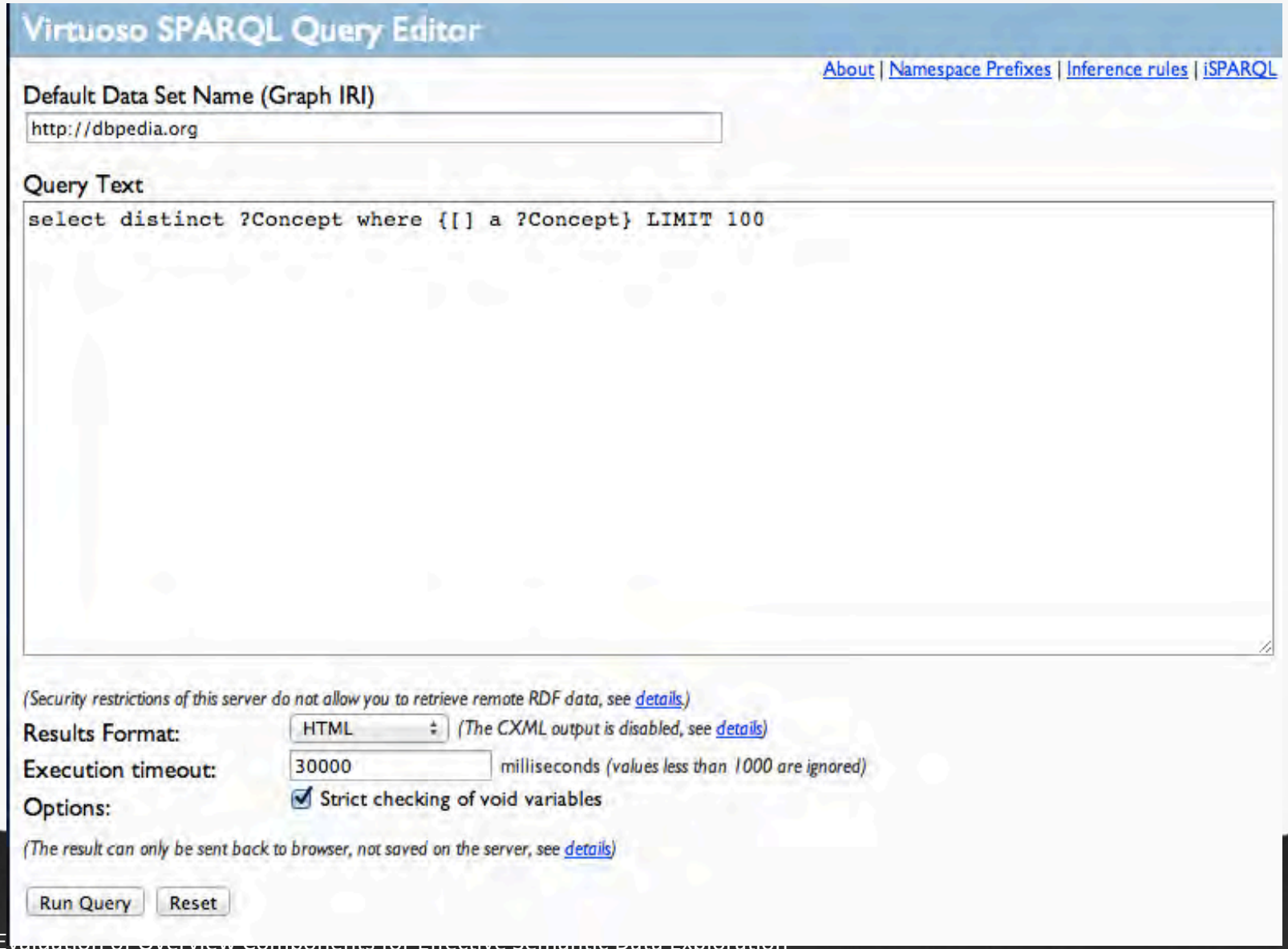
URI lookup


OPENLINK SOFTWARE
Text Search
Entity Label Lookup
Entity URI Lookup
Featured | Demo Queries | About
Precision Search & Find
URI
Describe
Hint: You can add this engine in search bar of an OpenSearch - capable browser
Faceted Search & Find service v1.11.15
POWERED BY VIRTUOSO
LINKINGOPENDATA
OpenLink Virtuoso version 07.00.3203, on Linux (i686-generic-linux-glibc212-64), Standard Edition
Data on this page is owned by its respective rights holders.
Virtuoso Faceted Browser Copyright © 2009-2012 OpenLink Software

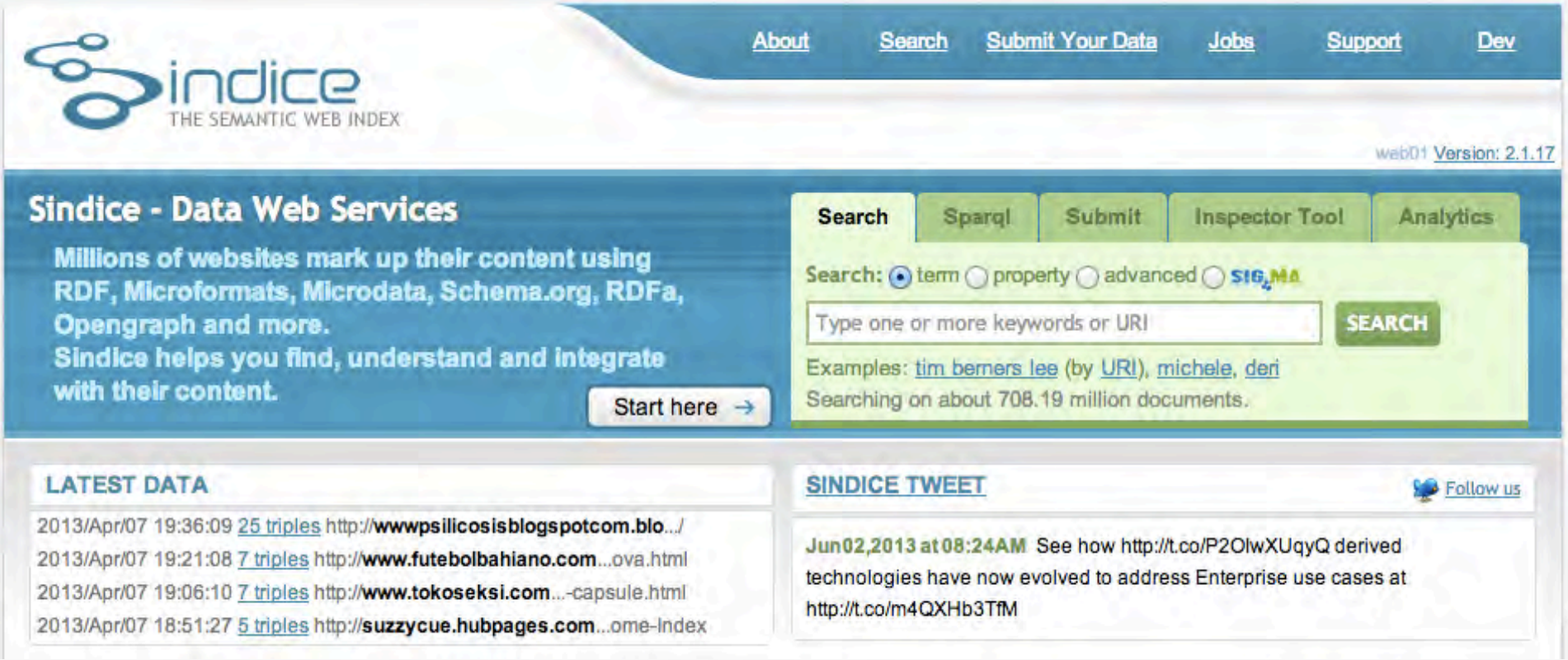
Graph Visualization



SPARQL


Virtuoso SPARQL Query Editor
About | Namespace Prefixes | Inference rules | iSPARQL
Default Data Set Name (Graph IRI)
http://dbpedia.org
Query Text
select distinct ?Concept where {[] a ?Concept} LIMIT 100
(Security restrictions of this server do not allow you to retrieve remote RDF data, see details.)
Results Format:
HTML
(The CXML output is disabled, see details)
Execution timeout:
30000
milliseconds (values less than 1000 are ignored)
Options:
Strict checking of void variables
(The result can only be sent back to browser, not saved on the server, see details)
Run Query
Reset

Keyword search



Human Semantic Web Interaction

- **Semantic Web HCI Challenges:**
  - Lot of data but... available as RDF data dumps, semantic query services, etc
  - Text search, type URI, SPARQL query,…
    …but they do not answer **end-users needs**
  - It requires **Semantic Web expertise**
  - Existing tools **hardly usable** for **lay users**
  - **No overview**

Approach

- **Visual Information-Seeking Mantra** for data analysis [Shneidermann]

"**Overview first,**
zoom and filter,
then details-on-demand"

Approach

- **Overview Task:**
  - **First step** to explore a dataset
  - Appreciation of the **size** and **extent** of the data
    - What is the data about? What is **not** the data about?
  - Serves as a **starting point for navigation**

- **Overview in SW data:**
  - **Main classes** and number of instances
  - Class **hierarchy**

Proposal

- Adapt existing **Information Architecture components** to Semantic Data scenarios to provide this **Overview**:
  - Navigation menus
  - Site map
  - Site index
  - Treemap (visualization technique)

- **Automatically generated** from the dataset structure and ontologies

Navigation Menus

- Overview of the **main classes**
- **Consistent navigation**: present in every page
- Automatic generation:
  - Parameters:
    - Flatten to desired **number of levels**
    - **Elements per level**
  - Algorithm:
    - When there is room → **divide class** with most instances
    - When too many options → **group classes** with less instances

Navigation Menus

Site Map

- **Overview** of the content at a **single glance**
- Main classes organized **hierarchically**
- Two versions:
  1. **Summarized version** related with Navigation Menus (only two levels)
  2. **Full version** related with the original hierarchy

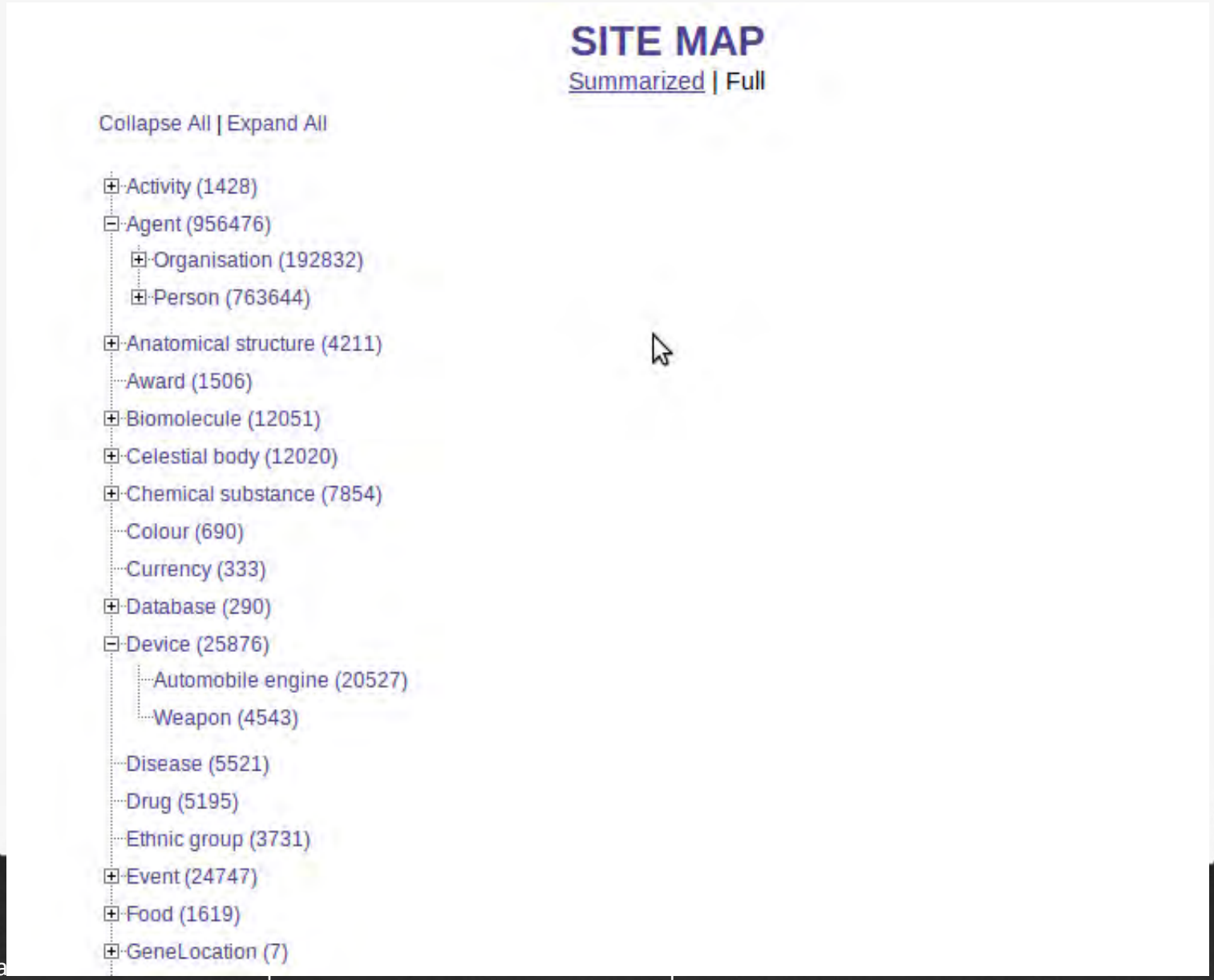
Summarized Site Map

SITE MAP

Summarized | Full

- Agent (2652224)
  Artist (116355), Athlete (371886), Cleric (14570), Fictional character (14892), Military person (21755), Office holder (32373), Organisation (475799), Organisation member (554952), Politician (38491), Ambassador (369), Architect (1186), Astronaut (621), College coach (5271), Criminal (1426), Journalist (10), Judge (1536), Model (1257), Monarch (2165), Philosopher (1198), Playboy Playmate (935), Poker player (615), Royalty (13126), Scientist (12842), Soccer manager (12118),
- Mean of transportation (139932)
  Aircraft (8329), Automobile (7486), Locomotive (2655), Rocket (233), Ship (51128), Space shuttle (20), Space station (34), Spacecraft (81),
- Person function (90488)
- Place (2010705)
  Architectural structure (343197), Historic place (6426), Monument (4), Natural place (158363), Populated place (920759), Protected area (6878), Site of Special Scientific Interest (461), Ski area (524), Wine region (321), World Heritage Site (1044),
- Species (402250)
  Archaea (171), Bacteria (328), Bird (12401), Fish (14291), Fungus (8244), Insect (62436), Mammal (8598), Mollusca (26282), Plant (49658), Amphibian (6892), Arachnid (2983), Crustacean (2212), Reptile (5415),
- Work (738684)
  Album (112249), Film (71715), Musical (1146), Single (41774), Software (46291), Song (6139), Television episode (7290), Television show (23480), Website (2605), Written work (92725),
- Other (222635)
  Anatomical structure (6638), Biomolecule (24111), Celestial body (24040), Chemical substance (15708), Device (50946), Disease (5521), Event (54764), Language (6860), Sales (5213), Activity (2856), Award (1506), Colour (690), Currency (333), Database (580), Drug (5195), Ethnic group (3731), Food (2273), GeneLocation (14), Holiday (524), Legal Case (4666), Music genre (714), Name (4404), Olympic result (744), Programming language (470), Project (32), Snooker world ranking (102),

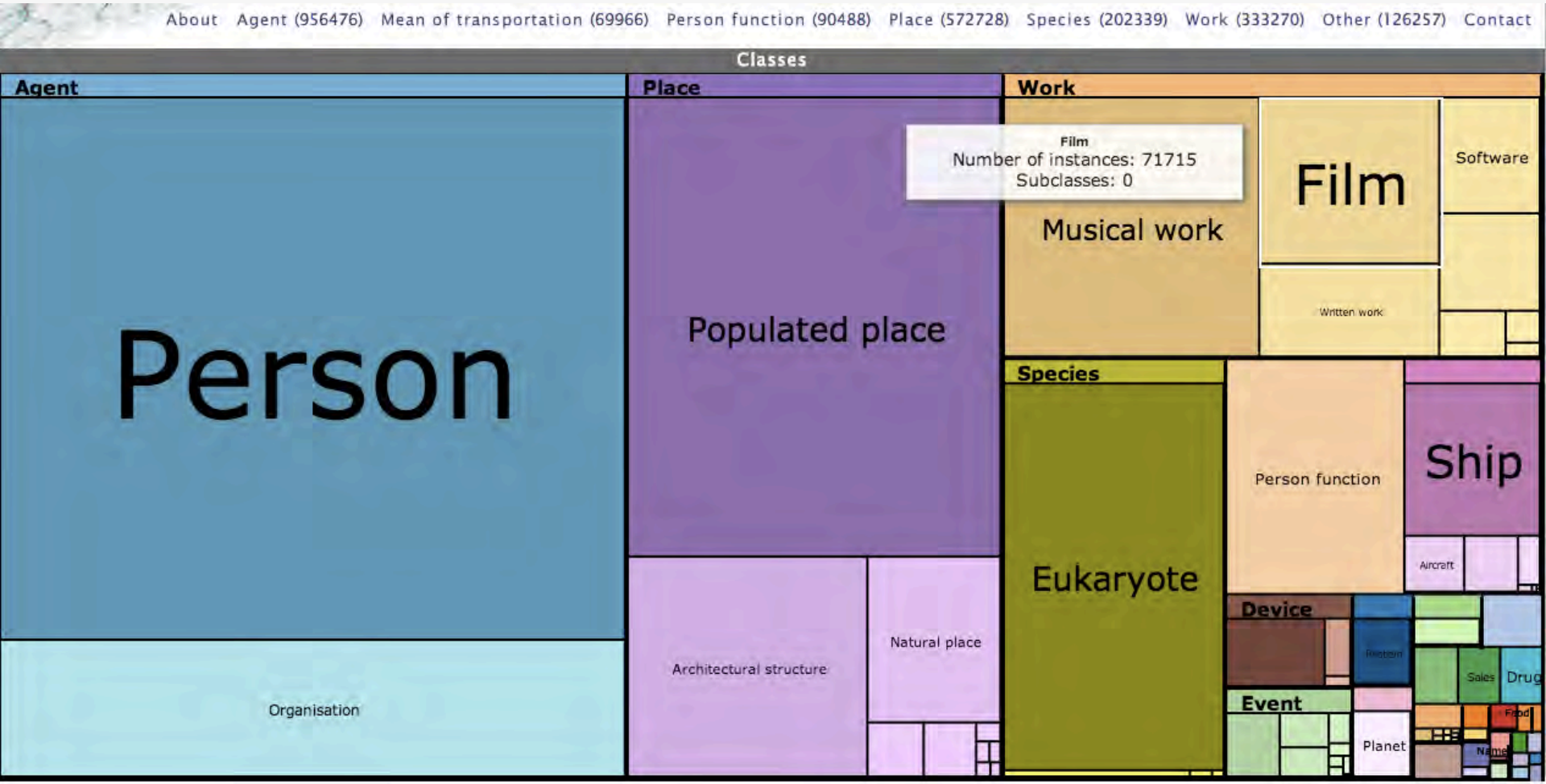
Full Site Map



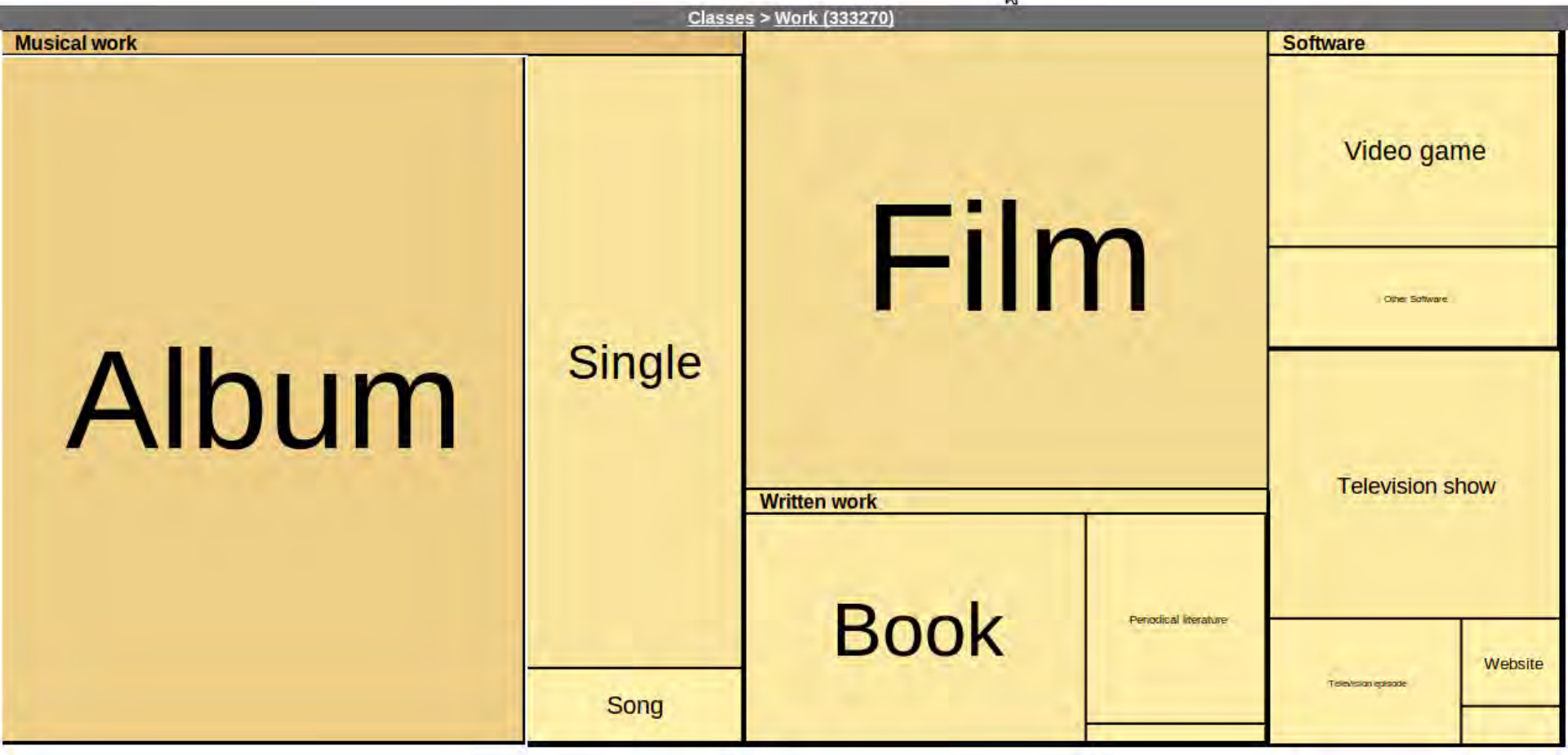
Treemap

- Visualization technique for **displaying hierarchical data**
- **Design:**
  - **Different colors** to group subclasses
  - Label **size proportional**
  - Show **two levels** + zoom in

Treemap


About
Agent (956476)
Mean of transportation (69966)
Person function (90488)
Place (572728)
Species (202339)
Work (333270)
Other (126257)
Contact
Classes
Agent
Person
Organisation
Place
Populated place
Architectural structure
Natural place
Work
Film
Number of instances: 71715
Subclasses: 0
Musical work
Film
Software
Written work
Species
Eukaryote
Person function
Ship
Aircraft
Device
Event
Planet
Sales
Drug
Food

Treemap


Classes > Work (333270)
Musical work
Album
Single
Song
Film
Written work
Book
Periodical literature
Software
Video game
Other Software
Television show
Television episode
Website

Site index

- List all pages/categories **alphabetically**
- Access to **particular content**
- Useful to search for a **specific page**

Site index


SITE INDEX
A | B | C | D | E | F | G | H | I | J | K | L | M | N | Ñ | O | P | Q | R | S | T | U | V | W | X | Y | Z |
A
Academic journal (4636)
Activity (2856)
Actor (4019)
Administrative region (32733)
Adult (pornographic) actor (1588)
Agent (3415868)
Aircraft (8329)
Airline (2984)
Airport (11533)
Album (112249)
Ambassador (369)
American football league (69)
American football player (11297)
American football team (3)
Amphibian (6892)
Anatomical structure (6638)
Animal (287419)
Arachnid (2983)

Evaluation

- Tests conductes at the **UsabiliLAB:**
  - Morae Recorder + Morae Observer
- **10 participants** (6 males, 4 females)
  - good IT knowledge
  - no SW expertise
  - never used Dbpedia before

Evaluation

- **Metrics:**
  - Effectiveness:
    - Completed
    - Partially completed
    - Not completed
  - Task time
  - User satisfaction

Evaluation - Tasks

- Tasks for the **common information needs:**
  - Known-item seeking:
  - Exhaustive research
  - Re-finding

Evaluation - Tasks

- **Known-item seeking**.
  Find **concrete classes** for different levels in the hierarchy
  - Task 1: find 1st level class – Work, Place, Event…
  - Task 2: find 2nd level class – Monument, Insect…
  - Task 3: find 3rd level class – Reptile, Sports team…
- Components:
  1. Navigation menus
  2. Site map
  3. Treemap
  4. Site index

Evaluation - Tasks

- **Exhaustive research:**
  - Task 4: comparing node size.
    Arrange from bigger to smaller:
    - Place, Agent and Species (1)
    - Work, Place, Mean of transportation (2)
    - Eukaryote, Person, Musical Work (3)
- Components:
  1. Navigation menus
  2. Site map
  3. Treemap

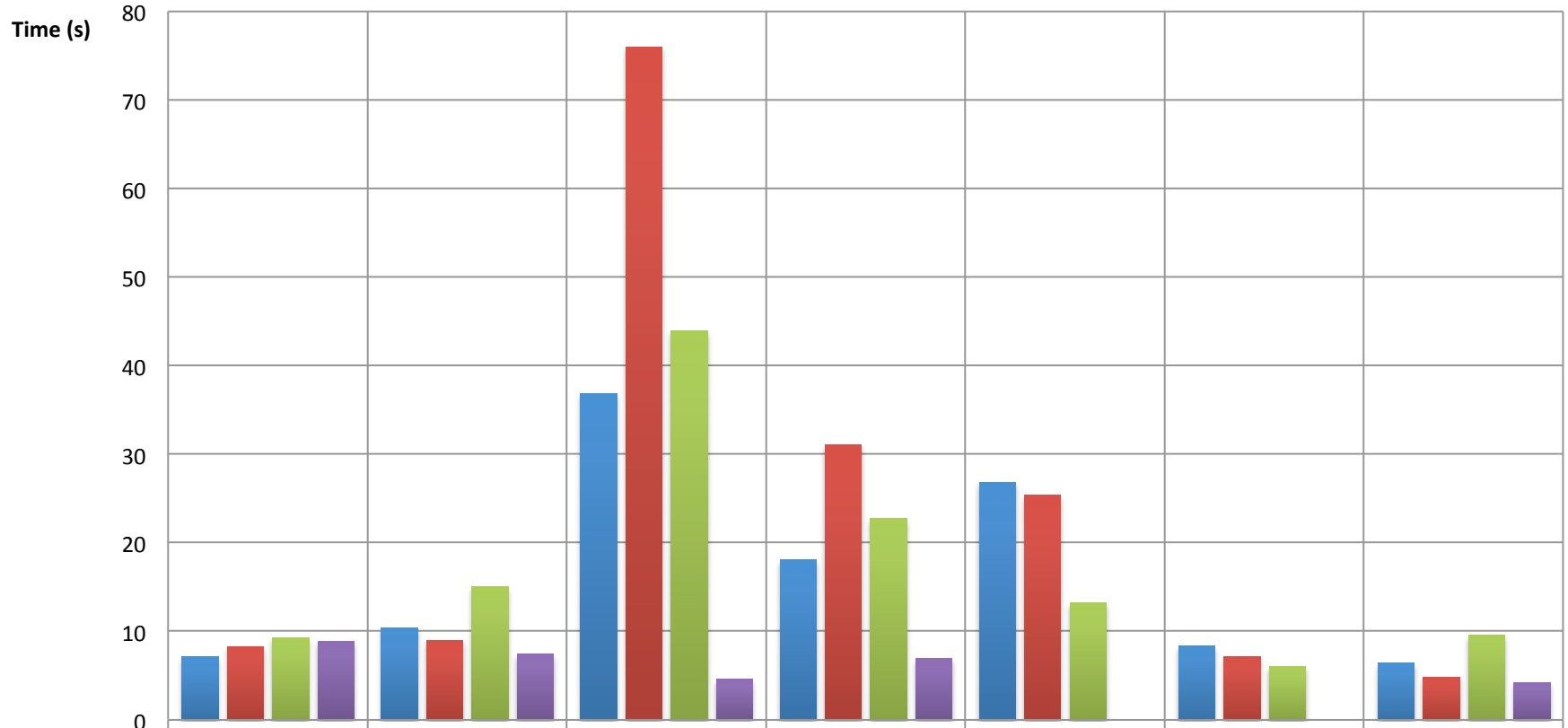
Evaluation - Tasks

- **Exhaustive research:**
  - Task 5: topology understanding.
    Find subclasses of:
    - Mean of transportation (1)
    - Species (2)
    - Work (3)
- Components:
  1. Navigation menus
  2. Site map
  3. Treemap

Evaluation - Tasks

- **Re-finding**.
  - Task 6: Find previously visited classes
    - Monument (1)
    - Insect (2)
    - Aircraft (3)
    - Film festival (4)
- Components:
  1. Navigation menus
  2. Site map
  3. Treemap
  4. Site index

Evaluation - Results



	T1. Known item 1st	T2. Known item 2nd	T3. Known item 3rd	Average T1, T2, T3	T4. Compare node size	T5. Find subclasses	T6. Re-finding
Navigation Menus	7,13	10,38	36,86	18,12	26,84	8,38	6,435
Site map	8,3	8,93	76	31,07	25,39	7,15	4,85
Treemap	9,3	15,03	43,97	22,76	13,23	6,04	9,55
Site Index	8,9	7,43	4,62	6,98			4,25

Evaluation - Satisfaction

Question (rate from 1 to 7)	NM	SM	TR	SI
Q1. The information showed on this component is helpful	6.2	6.2	5.7	6.4
Q2. This component is easy to use	6.4	6.2	5.7	6.6
Q3. Using this component, it was easy to find the information I was looking for	6.3	5.3	5.5	7
Q4. From this component, I understand the structure of the website	6.3	6.6	5.4	4.6
Q5. From this component, I understand what content is available on the website	6.2	6.3	4.9	5.8
Average	**6.28**	6.12	**5.44**	6.08

Conclusions

- Summary:
  - **4 components** to get an overview
  - **Automatically** generated
  - **Scalable**

- Make published semantic data more **usable** and **accessible** (no extra tools required, just browser)

Conclusions

- **Navigation menus:**
  - useful to find the **main important classes**
  - **limit number** of options
- **Site map**
  - general overall top-down view
  - **need to understand** the structure to find classes

Conclusions

- **Treemap:**
  - compare node sizes, identify groups by colors
  - **difficult** to find small nodes
- **Site index**
  - easy access to particular content
  - **no information** about structure

Future Work

- Improve components and usability
- Develop **keyword search**
- Evaluate with other datasets:
  - Different sizes
  - Different structures
- Overview for datasets **without schema**
  - **Clustering**

Thanks for your attention

Josep Maria Brunetti
josepmbrunetti@diei.udl.cat

http://rhizomik.net