



ontotext

Semantic Technology Lab



Massive Semantic Annotation and High-Performance Semantic Repositories

Atanas Kiryakov*

An Expedition to European Digital Cultural Heritage
Collecting, Connecting - and Conserving ?

Int. Conference, Residenz zu Salzburg

*presented by Wernher Behrendt, Salzburg Research

21 June, 2006

Outline

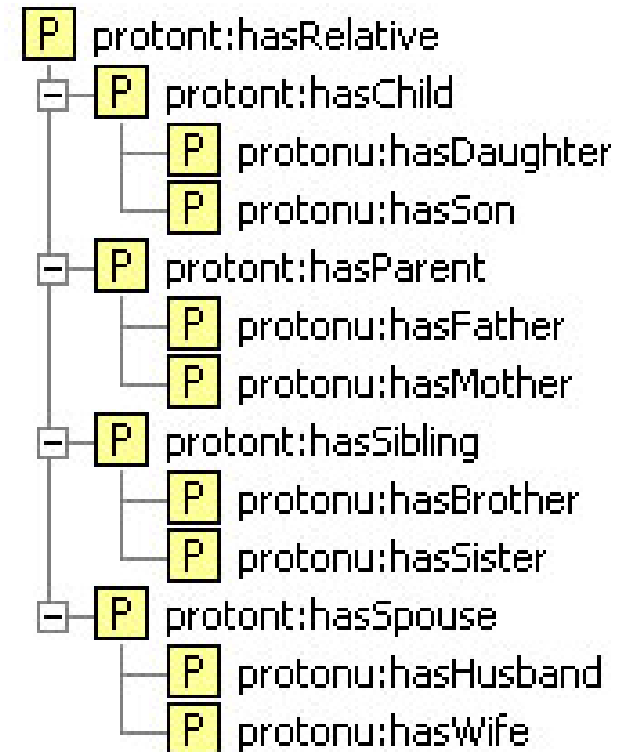
- **Ontologies and Knowledge Representation**
 - Semantic Repositories: [Databases plus machine reasoning!](#)
- **Semantic Web**
 - Connecting Diverse Data with Semantic Web: [Better interoperation!](#)
- **OWLIM – High Performance Semantic Repository**
 - Language Support for RDF and OWL ([Languages of the sem. Web!](#))
 - Scalability
- **KIM – Automatic Semantic Annotation**
 - Semantic Annotation ([what is it, why do it ?](#))
 - Hyperlinking, Highlighting
 - Co-Occurrence Analysis, Search and Timelines

Semantic Repositories

- **Semantic repositories** allow for
 - management and integration of heterogeneous data
 - Dynamic, easy, and flexible interpretation of multiple schemata
 - support Inference
- They can **replace RDBMS** in a wide range of applications
 - Suitable for analytical tasks and Business Intelligence (OLAP)
 - Usually suboptimal for highly dynamic transaction-oriented environments (OLTP)
- Semantic Repositories are for the Semantic Web what the HTTP servers were in the early days
 - But they are also useful for many other applications: lifesciences, business ontologies

Motivating Scenario for Ontologies

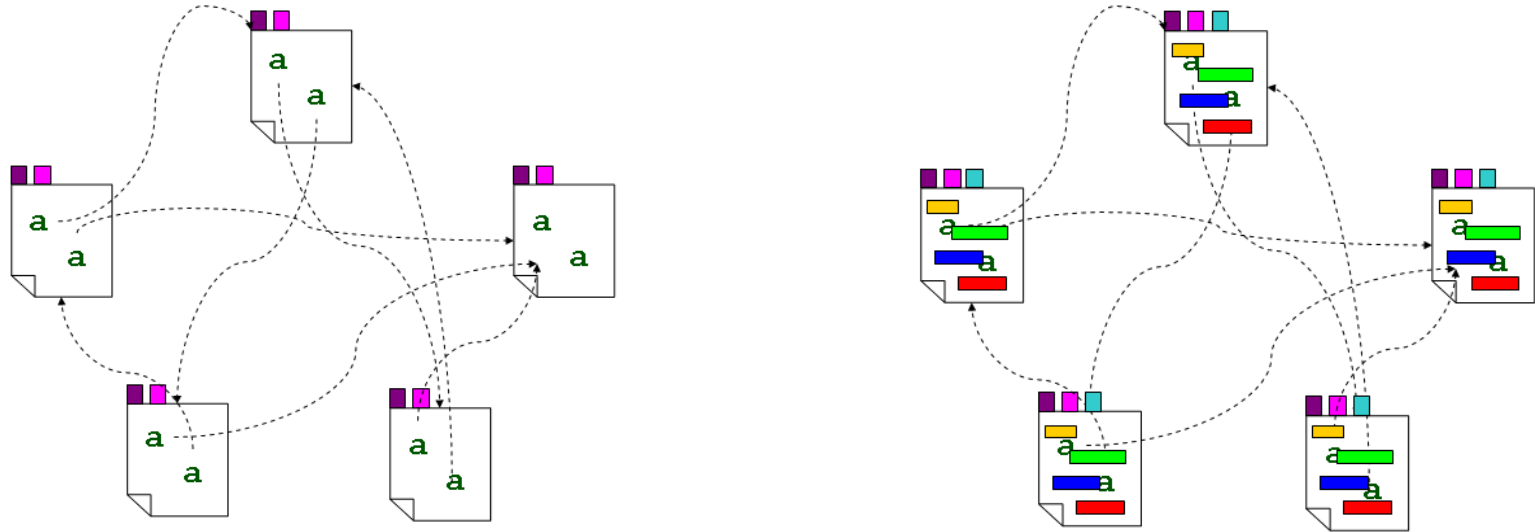
- In a **typical DB**:
 - If it was asserted that "*John is a son of Mary*"
 - Can answer to just a couple of questions:
 - *Who are the son(s) of Mary?*
 - *Of whom is John the son?*
- In contrast, a **semantic repository**:
 - Given simple family-relationships ontology
 - It can infer (and answer) the more general fact that John is a child of Mary (because *hasSon* is a sub-property of *hasChild*)
 - Mary and John are relatives (in both directions, because *hasRelative* is defined symmetric).
 - If it is known that Mary is a woman, it will infer that Mary is the mother of John – a more specific inverse relation



Outline

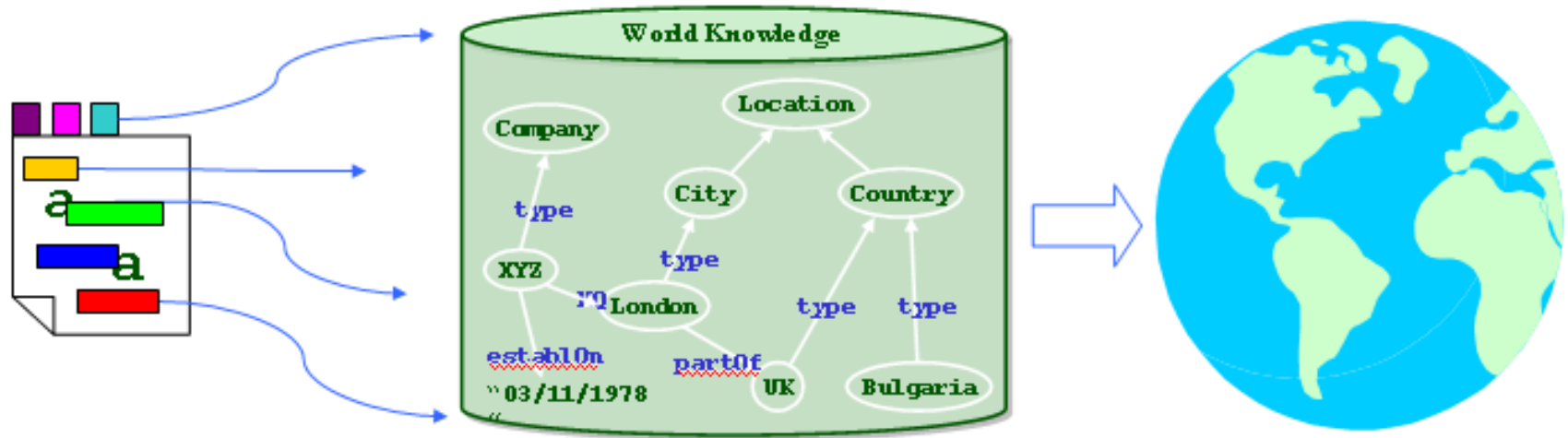
- **Ontologies and Knowledge Representation**
 - Semantic Repositories: [Databases plus machine reasoning!](#)
- **Semantic Web**
 - Connecting Diverse Data with Semantic Web: [Better interoperation!](#)
 - Languages for (semantic) interoperation: XML, RDF(S), OWL ...
- **OWLIM – High Performance Semantic Repository**
 - Language Support for RDF and OWL ([Languages of the sem. Web!](#))
 - Scalability
- **KIM – Automatic Semantic Annotation**
 - Semantic Annotation ([what is it, why do it ?](#))
 - Hyperlinking, Highlighting
 - Co-Occurrence Analysis, Search and Timelines

Semantic Web



- Semantic Web adds semantic metadata to the web
- “Semantic” metadata refer to machine-interpretable descriptions of the semantics of the documents or entities referred in them

Semantic Web (II)

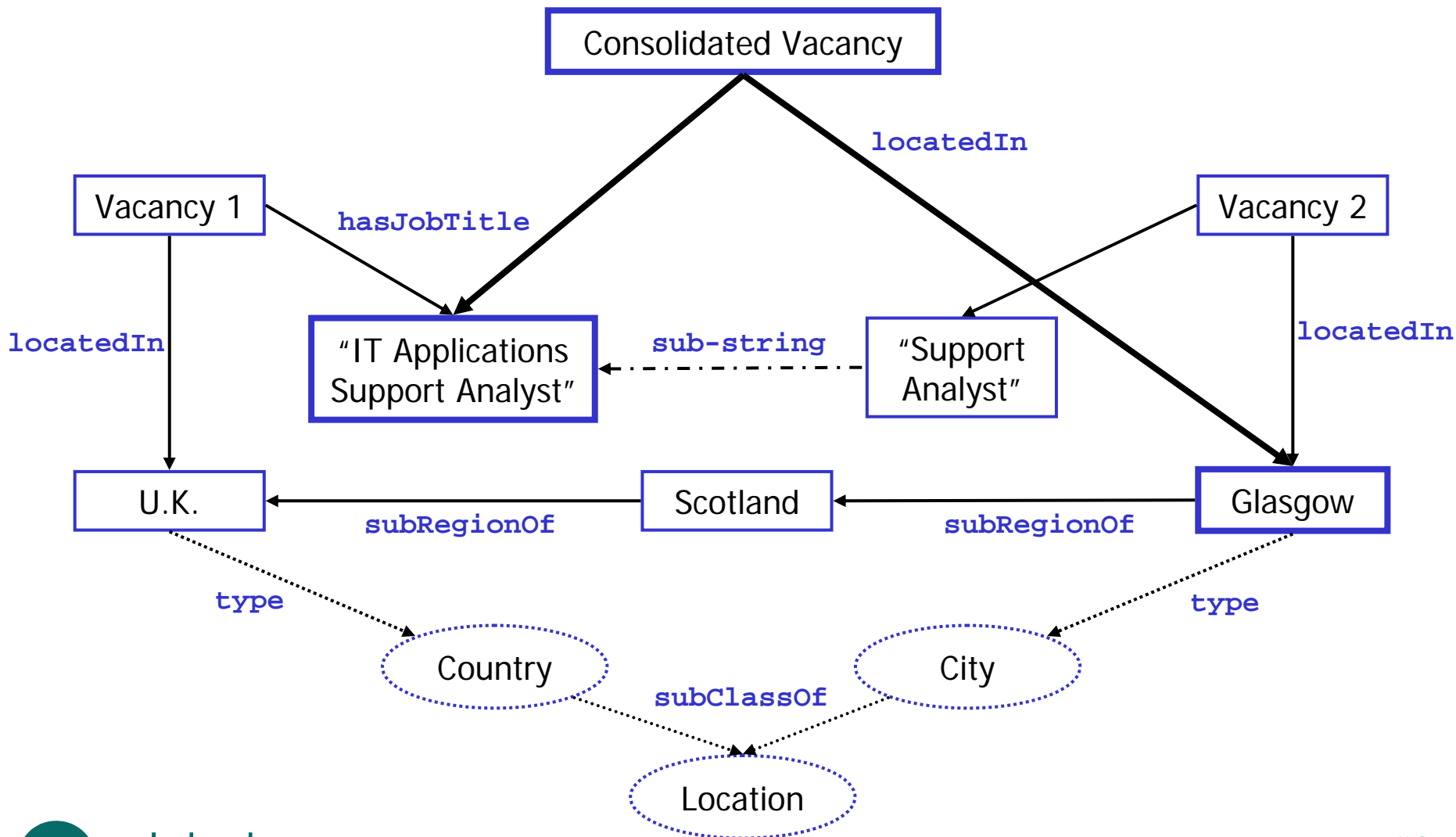


- Semantic Web adds semantic metadata to the web
- This metadata refer to machine-interpretable descriptions of the semantics of the documents or entities referred in them

Connecting Diverse Data with Semantic Web Technologies

- **Semantic Web standards** are designed for Web environment
 - No central authority
 - No consensus on the metadata schemata/ontologies
 - Several descriptions for one and the same object
- RDFS has the following features, critical for data integration:
 - **Globally unique identifiers** (URIs)
 - As compared to the local IDs in the databases
 - Open-world Assumption
(RDF does not care whether the world is open or closed)
 - Designed for complementary datasets
 - An RDF based-system can operate with contradictory data
(an RDF based system does not care whether the data is contradictory or not ...)

Example: Consolidation of Staff Vacancy Descriptions



Outline

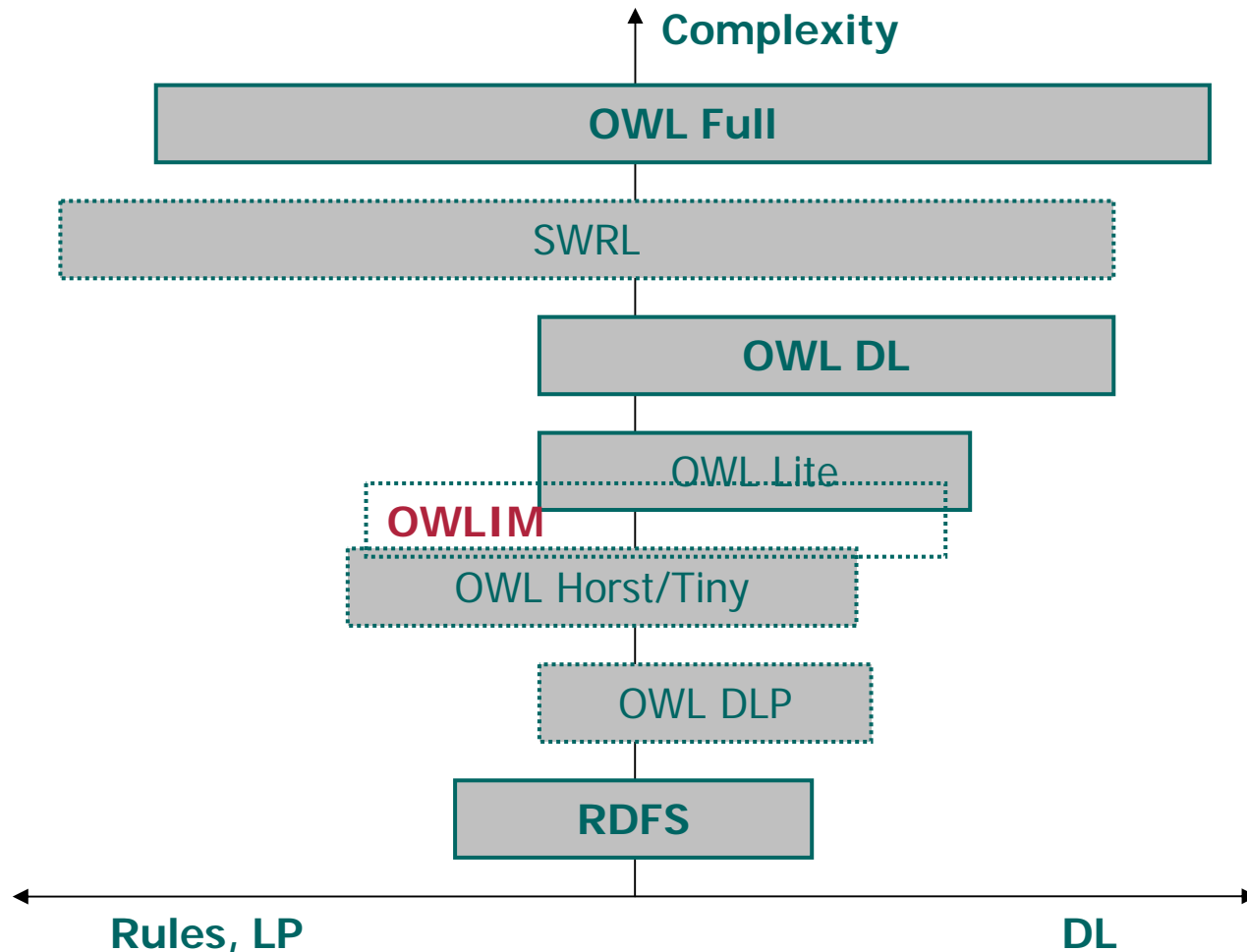
- Ontologies and Knowledge Representation
 - Semantic Repositories
- Semantic Web
 - Connecting Diverse Data with Semantic Web
- **OWLIM – High Performance Semantic Repository**
 - Language Support
 - Scalability
- KIM – Automatic Semantic Annotation
 - Semantic Annotation
 - Hyperlinking, Highlighting
 - Co-Occurrence Analysis, Search and Timelines

OWLIM: An OWL Semantic Repository (= RDF Database + machine reasoning)

- OWLIM is a **scalable semantic repository base on Sesame**
- OWLIM supports **full RDF(S) and limited OWL Lite**
 - It supports OWL Horst (more than OWL DLP) and rule-extensions
- OWLIM uses TRREE (Triple Reasoning and Rule Entailment Engine)
 - **forward-chaining** and “materialization”, “inferred closure” is maintained
- It performs **in-memory reasoning and query evaluation**
- Combined with **reliable persistence**, based on RDF N-Triples
- The compromise: **relatively slow delete** operation
 - “limited” scalability on scenarios with high implicit/explicit st. ratio
- **Very fast upload, retrieval, query evaluation** for huge KB

Naïve OWL Fragments Map

(The scientists are still arguing how complicated a knowledge representation language needs to be)



http://www.ontotext.com/inference/rdfs_rules_owl.html#owl_horst

Semantics Supported by OWLIM

(for researchers into RDF repositories)

- The reasoning support in OWLIM is customizable;
- The **rule set** parameter allows for switching between 4 predefined inference modes:
 - **owl-max** – the most expressive set (see the next slides);
 - **owl-horst** – a set similar to the one defined in [Horst05]:
 - It is sufficient to pass the LUBM benchmark correctly;
 - Similar to what was defined as OWL-Tiny at SWAD-Europe'03
 - **rdfs** – the standard RDF(S) semantics;
 - **empty** – as an RDF store without any inference;
- The **partRDFS** parameter allows switching on/off an optimization in the RDFS support (see slide: RDFS Support)

Language Support Comparison (for researchers into RDF repositories)

- The **owl-max** semantics of OWLIM is **close to OWL Horst**:
 - Owl-max is generally richer than the pD^* -entailment of Horst, but
 - OWLIM has no support for the inconsistency rules in pD^*
 - No datatype-supporting modifications to the RDFS semantics (D^*)
- OWLIM (owl-max) vs. OWL DLP
 - OWLIM supports a fragment **richer than OWL DLP**
 - OWLIM supports the full RDFS semantics, while DLP considers the DL-specific constraints
- OWLIM vs. OWL Lite:
 - OWLIM supports the full RDFS semantics, while OWL Lite does not
 - Incomplete support for some primitives

In-memory Reasoning and Reliable Persistence (machine reasoning by pre-computing the results)

- OWLIM uses Ontotext's TRREE (Triple Reasoning and Rule Entailment Engine)
 - TRREE implements **R-Entitlement**
 - for **forward-chaining** and “total materialization”
 - The “inferred closure” is generated and maintained up to date
- It performs **in-memory reasoning and query evaluation**
- Combined with **reliable persistence**, based on RDF N-Triples
- The compromise: **relatively slow delete** operation
 - “limited” scalability on scenarios with high implicit/explicit st. ratio
 - not the case with most of the popular ontologies in RDFS and OWL
- **Very fast upload, retrieval, query evaluation** for huge KB

A Configurable SAIL for Sesame (for RDF-Database techies)

- OWLIM is available as a Storage and Inference Layer (SAIL) for **Sesame RDF database** (v.1.2.1-1.2.4). Benefits:
 - Sesame's infrastructure, documentation, user community, etc.
 - Support for multiple query languages (RQL, RDQL, SeRQL)
 - Support for import and export formats (RDF/XML, N-Triples, N3)
 - Version compatible with Sesame 2.0 is almost ready
- **OWLIM Configuration Options:**
 - **noPersist**: the N-Triples persistency switched off
 - **Configurable semantics**: through the **ruleset** and **partialRDFS** parameters
 - **Configurable index size**: allows trading memory for performance
 - **stackSafe**: switches on a slower mode of the TRREE engine, which practically eliminates the possibility of stack overflow errors

OWLIM Default Configuration (for RDF-Database techies)

- **OWLIM's default configuration is:**
 - **noPersist=false** (i.e. it does store the content of the repository)
 - Rule-set/semantics: **owl-horst**
 - **partialRDFS=true**
 - Index-Size: **4M entries** (64MB of RAM; 16 bytes per entry)
 - **stackSafe=false**
- The configuration of OWLIM which provides the same functionality as the Sesame's most popular in-memory **RDFSSchemaSail** is:
 - **noPersist=true;**
 - Rule-set/semantics: **RDFS**
 - **partialRDFS=false**

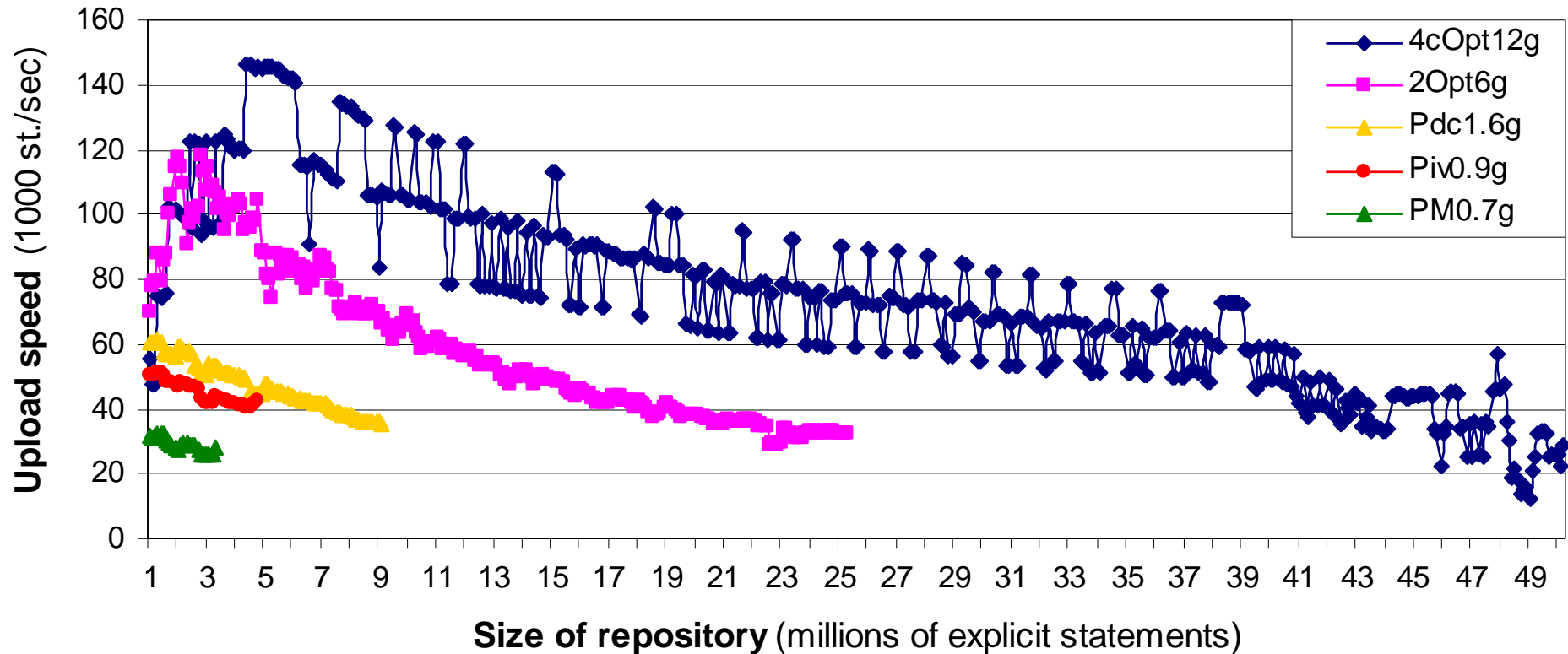
Performance Evaluation Configurations

Name	Configuration	RAM (jvm -Xmx)	JDK	Comment
4cOpt12g	2 x Opteron 270 (2.0GHz, dual-core), Suse Linux v.10, 64-bit	12GB, DDR400	JDK 1.5 64-bit	A DB/application server; SATA2 drives; RAID10; ~4000 EURO
2Opt6.0g	2 x Opteron 246 (2.0GHz) Windows Server 2003 64-bit	6GB, DDR400	JDK 1.5 64-bit	A DB/application server; SATA2 drives; RAID10; ~3000 EURO
Pdc1.6g	Pentium D 920 (2.8GHz, dual-core), Win XP	1.6GB, DDR2 667	JDK 1.5	Workstation
Piv0.9g	Pentium IV 630 (3.0GHz), Win XP	900MB, DDR2 533	JDK 1.5	Office desktop
Pm0.7g	Pentium Mobile 1.6GHz, Win XP	700MB, DDR266	JDK 1.5	Notebook (Q2'03)

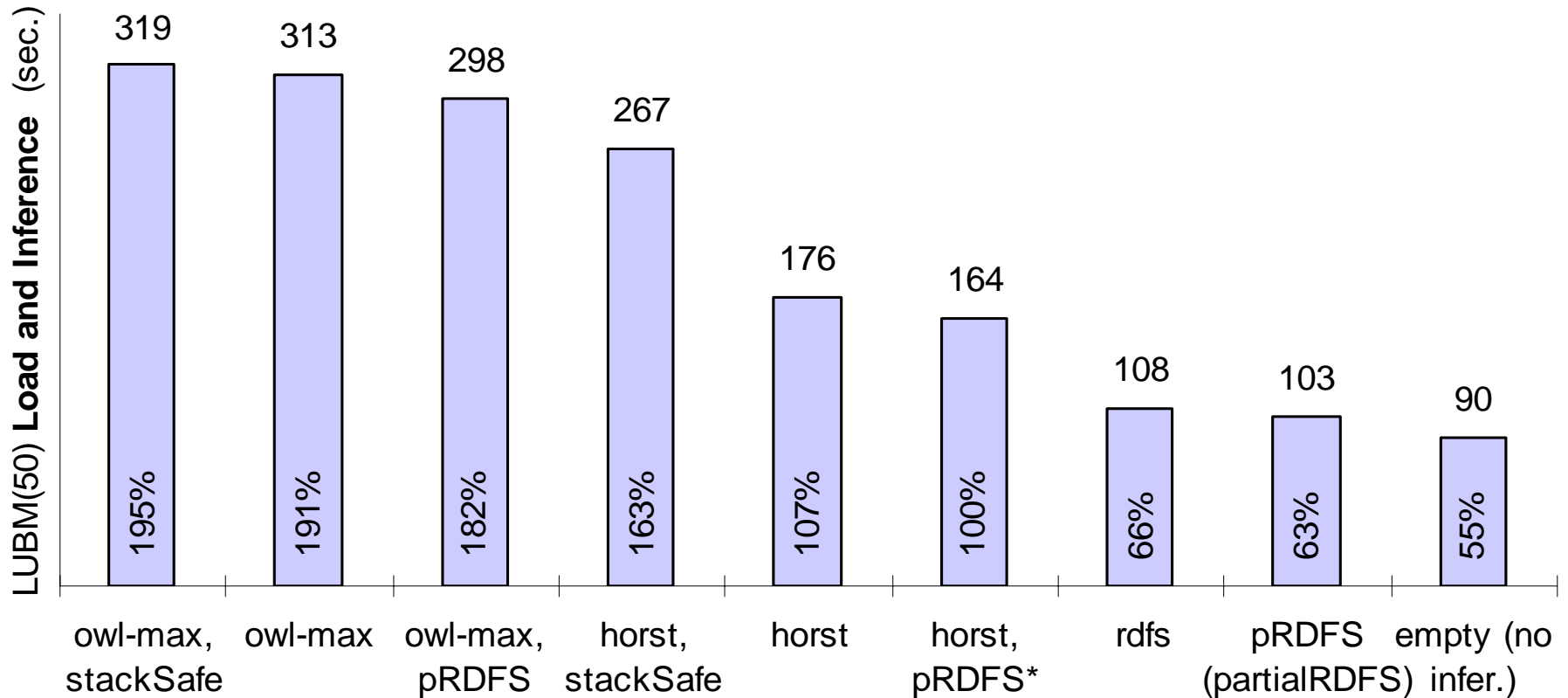
City Benchmark

- The repository is **pre-populated with about 500k explicit statements**:
 - a real ontology (PROTON) and
 - a knowledge base – the small version of KIM's World Knowledge Base;
- **Synthetic descriptions of cities** are added incrementally,
 - each transaction adds one city described in about **10k statements**.
- **Interlinking to the non-synthetic part**:
 - Ten synthetic organizations are created and “located” into each city;
 - 38 persons are created and settled within each of the organizations;
- A couple of test queries (in SeRQL) are evaluated after the addition of each 10 cities (i.e. after each new 100k statements).
- Delete gets roughly 10 sec. per million of statements

OWLIM Performance: Upload and Inference



LUBM(50,0): Rule-set and Inference Mode (Lehigh University Benchmark - <http://swat.cse.lehigh.edu/pubs/guo05a.pdf>)



LUBM(50,0): Rule-set and Inference Mode (II)

- The **partRDFS optimization provides 6-7% speedup**
- The **inference takes a bit less than half of the processing time**
 - with the default configuration: owl-horst with partRDFS
 - considering that the plain RDF version (empty) requires only 55% of the time to load the dataset;
- **owl-max is twice slower**
 - than the default setup owl-horst
 - In both cases partialRDFS switched on
- The **stack-safe mode slows down by 50%**
 - the version owl-horst with partialRDFS switched off
 - but has almost no impact on the owl-max inference

BigOWLIM

- **Fully functional pre-release** of BigOWLIM is already available:
 - Check <http://www.ontotext.com/owlim/big/>
- BigOWLIM is an even more scalable **not-in-memory** version, based on the corresponding version of the TRREE engine
 - The “standard” OWLIM version, which uses in-memory reasoning and query evaluation is referred as **SwiftOWLIM**
- BigOWLIM does not need to maintain all the contents of the repository in the main memory in order to operate
- BigOWLIM stores the contents of the repository (including the “inferred closure”) in binary files; not in N-Triples
 - This allows **instant startup and initialization** of large repositories, because it does not need to parse, re-load and re-infer all the knowledge from scratch

BigOWLIM vs. SwiftOWLIM

- BigOWLIM uses **sorted indices**
 - While the indices of SwiftOWLIM are essentially hash-tables
 - In addition to this BigOWLIM maintains data statistics, to allow ...
- Database-like **query optimizations**
 - Re-ordering of the constraints in the query has no impact on the execution time
 - Combined with other optimizations, this feature delivers dramatic improvements to the evaluation time of “heavy” queries
- Special handling of **equivalence classes** (owl:sameAs)
 - Large equivalent classes does not cause excessive generation of inferred statements

BigOWLIM: 100M Statements on a Desktop

- It handles **LUBM(1000,0) on desktop machine** with 2GB of RAM
 - Hardware: Piv0.9g (Pentium 4, 3.0GHz, #630)
 - 32-bit JDK 1.5 given -Xmx1600
 - **Loading, inference, and storage takes 11h 20 min.**
 - LUBM(1000,0) contains above **130M explicit statements**
 - 10GB RDF/XML files; the pure parsing time is about 4h
- **LUBM(50,0) processed with only 192MB** of RAM
 - Piv0.9g; 32-bit JDK 1.5 given -Xmx1600
 - **Loading, inference, and storage takes 26 min**
 - It is only 4 times slower than the in-memory version
 - Average upload speed around 4,000 st./sec.

Reasoning over 1 Billion Statements

- BigOWLIM successfully passed **LUBM(8000,0)**
 - Hardware: 4cOpt12g (2 x Opteron 270, 16GB of RAM, RAID 10)
 - OS: Suse 10.0 Linux, x86_64, Kernel 2.6.13-15-smp
 - 64-bit JDK 1.5 given -Xmx12000
 - **Loading, inference, and storage took 69 hours** and 51 min
 - LUBM(8000,0) contains **1.06 Billions of explicit statements**
 - The “inferred closure” contains about 786M statements
 - **Managing over 1.85 Billions of statements in total**
 - 92GB RDF/XML files; 95 GB binary storage files
 - **Average Speed: 4 538 statements/sec.**

Outline

- Ontologies and Knowledge Representation
 - Semantic Repositories
- Semantic Web
 - Connecting Diverse Data with SW Technologies
- OWLIM – High Performance Semantic Repository
 - Language Support
 - Scalability
- **KIM – Automatic Semantic Annotation**
 - Semantic Annotation
 - Hyperlinking, Highlighting
 - Co-Occurrence Analysis, Search and Timelines

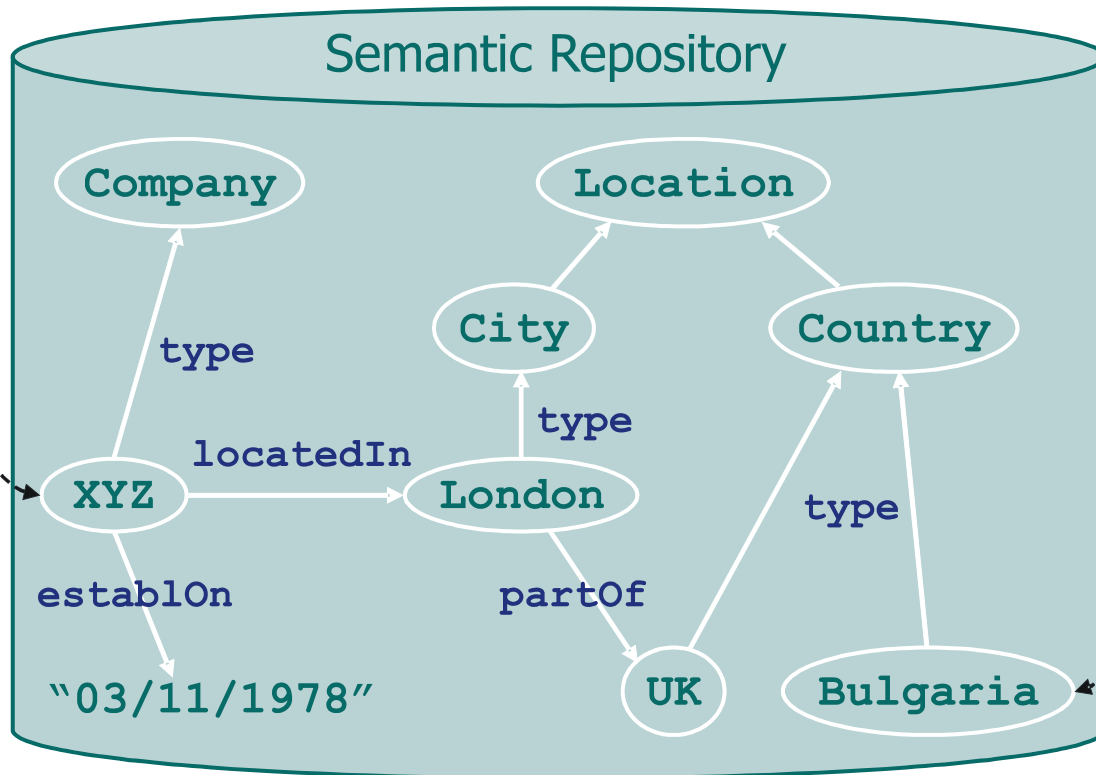
The KIM Platform

(Knowledge and Information Management)

- KIM is a software platform for:
- **Semantic annotation** of text
 - automatic ontology population
 - open-domain dynamic semantic annotation of unstructured and semi-structured content for Semantic Web and KM applications.
- Indexing and retrieval
- Query and exploration of formal knowledge.
- KIM includes:
 - PROTON, KIMSO, KIMLO and KIM World KB;
 - KIM Server – with API for remote access and integration;
 - Front-ends: KIM Web UI and Plug-in for Internet Explorer.
- A public KIM server is available at <http://ontodemo.sirma.bg/KIM>.

What KIM does? – Semantic Annotation

XYZ announced profits in Q3, planning to build a \$120M plant in Bulgaria, and more and more and more and more text...



Simple Usage: Highlight, Hyperlink, and ...

Guardian Unlimited | World Latest | EPA Moving on New Front to Cut Pollution - Microsoft Internet Explorer

Address <http://www.guardian.co.uk/uslatest/story/0,1282,-4076961,00.html>

KIM Plugin

- Entity
 - Abstract
 - Happening
 - Event
 - Situation
 - TimeInterval
 - Object
 - Agent
 - Organization
 - Person
 - BusinessObject
 - InformationResource
 - Location
 - Statement
 - Vehicle

Classes Entities Config

Place Links

Breaking news US

EPA Moving on New Front to Cut Pollution

Tuesday May 11, 2004 7:46 AM

By H. JOSEF [HEBERT](#)

[Associated Press](#) Writer

[WASHINGTON](#) ([AP](#)) - The government is moving on a new front to cut air pollution. This time ferry boats and harbor tugs, farm tractors and train locomotives, and dirt movers at construction sites are the targets.

The [Environmental Protection Agency](#) is issuing new regulations aimed at cutting the amount of smog-causing chemicals and fine soot that comes from these off-road diesel-powered vehicles

[Cheney to Have Routine Check of Pacemaker](#)
8:46 am

[GOP Seeks to End Tax Cut Debate](#)
8:46 am

[Govt. Grounds](#)

Simple Usage: ... Explore and Navigate

The screenshot shows two overlapping windows from Microsoft Internet Explorer. The left window, titled 'KIM Explorer - Microsoft Internet Explorer', displays a semantic interface for 'The Associated Press, a NewsAgency, Trusted^{tip!}'. It features two tables: one for properties and values, and another for related entities with links to 'The Associated Press'. The right window shows a news article with several terms highlighted in colored boxes: 'May 11, 2004 7:46 AM' (green), 'SEF HEBERT' (pink), 'ed Press W' (purple), and 'GTON (AP)' (yellow). A large orange arrow points from the right window back to the KIM Explorer window, indicating navigation. The bottom of the KIM Explorer window shows a 'Classes' panel with checked items: BusinessObject, InformationResource, Location, Statement, and Vehicle. Below this are buttons for 'Classes', 'Entities', and 'Config', and a 'Place Links' checkbox.

Property	Value
hasAlias	The Associated Press
hasAlias	AP
hasAlias	Associated Press
locatedIn	New York
locatedIn	New York

Resource	Link to The Associated Press
correspondent	withinOrganization
correspondent	withinOrganization
correspondent	withinOrganization
correspondent	withinOrganization

Copyright © 2004 Ontotext Lab, Sirma AI, Bulgaria

Classes: BusinessObject, InformationResource, Location, Statement, Vehicle

Buttons: Classes, Entities, Config

Place Links

Simple Usage: ... Enjoy a Hyperbolic Tree View

The screenshot shows a Microsoft Internet Explorer window titled "TouchGraph GraphLayout - Microsoft Internet Explorer". The address bar contains the URL: `http://62.213.161.156/kim/graph/Graph.jsp?uri=http://www.ontotext.com/kim/kimo.rdfs%23PublicCompany_T.138`. The main content area displays a hyperbolic tree view centered on "Groupe Danone a trusted PublicCompany". The tree structure is as follows:

- Groupe Danone a trusted PublicCompany** (Central Node)
 - Jacques Vincent** (hasPosition, holder)
 - Vice Chairman and COO** (withinOrganization)
 - Franck Riboud** (hasPosition, holder)
 - Chairman and CEO** (withinOrganization)
 - Food, Beverage & Tobacco** (activeInSector)
 - French Republic** (locatedIn)
 - Danone** (fullyOwns)
 - New York Stock Exchange** (tradedOn)
 - http://www.danonegroup.com** (hasWebPage)
 - EVP, Finance** (withinOrganization)

On the right side, there is a panel titled "Attributes of Groupe Danone:" with the following information:

- Relations From: 5
- Relations To: 3
- hasMainAlias - Groupe Danone
- stockExchangeIndex - DA
- FISCAL_SALES - \$12,897 mln.
- FISCAL_NET_INCOME - \$118 mln.
- numberOfEmployees - 100,560
- comment - You say Danone, I say Danone. Let's call the whole thing one of the largest food producers in the world. Groupe Danone is the global leader in cultured dairy products (including yogurt, cheese, and dairy desserts) and biscuits (cookies, crackers, and snacks). Its Evian and other brands make it #2 in bottled water (behind Nestle).

Below the attributes is a "History:" section listing:

- Groupe Danone
- EVP, Finance
- Emmanuel Faber

At the bottom of the browser window, there are controls for "Zoom", "Rotate", and "Hyperbolic" (selected), along with the text "Use horizontal scrollbar to zoom and rotate the graph".

How KIM Searches Better

KIM can match a **query like**:

Documents about a **telecom company** in **Europe**,
John Smith, and a date in the **first half of 2002**.

With a document containing:

"At its meeting on the **10th of May**, the board of **Vodafone**
appointed **John G. Smith** as CTO"

The semantic IR can match:

- **Vodafone** with a "**telecom in Europe**", because:
 - Vodafone is a **mobile operator**, which is a **sort of telecom**;
 - Vodafone is **in the UK**, which is a **part of Europe**.
- **5th of May** with a "date in first half of 2002";
- "John G. Smith" with "John Smith".

Outline

- Ontologies and Knowledge Representation
 - Semantic Repositories
- OWLIM – High Performance Semantic Repository
 - Language Support
 - Scalability
- Semantic Web
 - Connecting Diverse Data with Semantic Web
- KIM – Automatic Semantic Annotation
 - Semantic Annotation
 - Hyperlinking, Highlighting
 - **Co-Occurrence Analysis, Search and Timelines**

CoreDB: Name and Goals

CoreDB is a component of KIM (Co-Occurrence and Ranking of Entities DB)

- **Number of appearances** and **popularity** of entities
Q1: How often has a **company** **appeared** in the international business news during a **given period** ?
- **Co-occurrence** of entities
Q2: Give me the people that **co-appear** with telecom **companies**.
- Combination of these with **semantic queries** and restrictions
Q3: Q2 + where the documents contain "fraud" and the company is **located in South-east Europe**

The Scale of Ambition

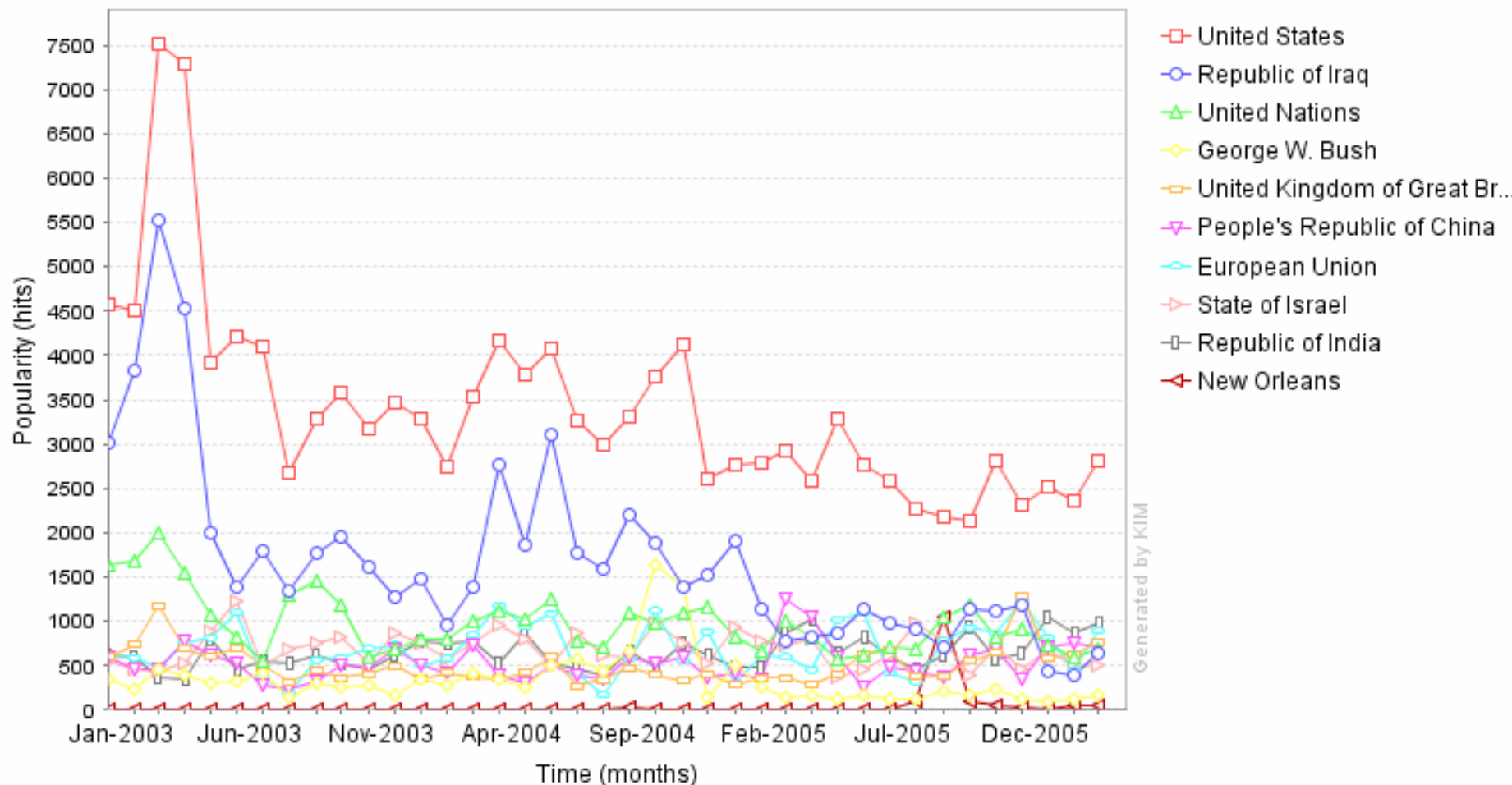
- The major point is to allow such queries in *efficient* manner over data with the following cardinality:
 - 10^6 entities/terms
 - 10^7 documents
 - 10^2 entities occurring in an average document
- This means managing and querying efficiently 10^9 entity occurrences
- We had tested the current implementations with 10^7 occurrences and it answers the basic queries in milliseconds.

CoreDB Applications

- **Detection of “associative” links** between entities, based on co-occurrence in documents
 - It is an alternative of the detection of strong links based on local context parsing
- **Ranking, measuring popularity of an entity** over a set of documents
 - The ranking is as good/relevant/representative as the set of documents is
- **Computing timelines** (changes over time) for entity ranking or co-occurrence
 - “How did our popularity in the IT press change during June” (i.e. “What is the effect of this 1.5MEuro media campaign ?!?”)
 - “How does the strength of association between organization X and RDF change over Q1 ?”

Timelines Result

Timelines for Most popular Entities



Document Filter: ALL docs, containing Keywords: (none) Entities: (none)

Time Period: 01/01/2003 to 31/03/2006 Granularity: Month

Options: display 2 topmost entities of type Entity for each time unit


CORE Search

| [Clear](#) | [Options](#) |

KIM

- Home
- Entity Pattern Search
- Predefined Patterns
- Entity Lookup
- Keyword Search
- Browse Ontology
- CORE Search
- Timelines
- About KIM

Powered by:



Document Keyword Filter

Matching documents: **63963**

[Documents](#) [Timelines](#)

Selected Items Filter

(No items selected)

Recent Items

(No recent items)

Key Phrases

25 of **14488** shown below

- abortion
- bloc
- chair
- church
- crowd
- earthquake
- election
- Faith
- freedom
- influence
- island
- Mass
- member nation
- papacy
- people
- Pope
- post
- quake
- reign
- scale
- sense
- summit
- trip
- Visit
- World

People

25 of **63161** shown below

- A. Tarzartes
- Bruce L.A. Carter
- Craig D. Shimasaki
- D. Gregory Smith
- Frank Chang
- Fred Hiller
- Fritz Horlacher
- George B. Rathmann
- Horace J. Davis III
- Ian Kindred
- J. Bruce Robinson
- James A. Johnson
- James J. Jim Schiro
- James R. Tolbert III
- Julie A. Sansom-Reese
- Lodewijk Christiaan van
- Makoto Kaneko
- Markus Haefeli
- Norman R. Proulx
- Patrick O'Sullivan
- Paul F. Forman
- Peter Eckert
- Richard M. Dressler
- Roberto R. Romulo
- William Kim Wah Tan

Organizations

25 of **27622** shown below

- Amhrest College
- Brandeis University
- Brigham Young Universit
- Drexel University
- Nanyang Technological L
- New Mexico State Unive
- Queen Marry University
- Sofia University
- Stanford University
- Technical University Kos
- TU Vienna
- University of Arizona
- University of Economics,
- University of Georgia
- University of Maryland
- University of Maryland C
- University of Santiago de
- University of South Calif
- University of Southern C
- University of Strathclyde
- University of Twente
- University of Washingtor
- Vassar College
- Vrije Universiteit Amsterc
- Vrije Universiteit, Brusse

Locations

25 of **21568** shown below


- Bandirma Korfezi
- Bristol Channel
- Dzharylhats'ka Zatoka
- Erdek Korfezi
- Feodosiys'ka Zatoka
- Garabogazkol Aylagy
- Gemlik Korfezi
- Golfo de Venezuela
- Golfo Triste
- Gulf of Aden
- Gulf of Antalya
- Gulf of Iskenderun
- Gulf of Paria
- Ildir Korfezi
- Izmit Korfezi
- Karkinit's'ka Zatoka
- Kusadasi Korfezi
- Persian Gulf
- Saros Korfezi
- Taganrogskiy Zaliv
- Tendrivs'ka Zatoka
- Turkmenbasy Aylagy
- Yahorlyts'kyy Lyman
- Zaliv Adzhibay
- Zatoka Syvash

Copyright © 2005 Ontotext Lab, Sima Group Corp.

Co-occurring Entities

CORE Search

| Clear | Options |

Document Keyword Filter	Key Phrases	People	Organizations	Locations
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Matching documents: 13 <input type="button" value="Documents"/> <input type="button" value="Timelines"/>	25 of 42 shown below	Loading ...	19 matching entities	25 of 49 shown below
Selected Items Filter <input type="checkbox"/> Graz <input type="checkbox"/> Republic of Austria	Bird case city coalition death governor honour judge name news agency newspaper paper parliament picture reform republic reputation resident ring son spokesman strike town union virus	Adolf Hitler Archbishop Rowan Willie  <u>Arnold Schwarzenegger</u> Axel Springer Verlag Benita Brian L Christian Fraser Christopher D. Hughes Deborah Lipstadt Donald Beardslee Greville Janner Irving Joerg Haider Markos Kyprianou Matthew Cooper Mozart Otto Schneider Pilz Quentin State Prison Rudolf Gollia Samuel Jutzi Siegfried Nagl Stanley Stephen Smith	AFP Austrian TV BBC BBC News Court of Appeals Green Party Holocaust Educational Tr Mediaprint National Federation of A Reuters Group PLC Robert Koch Institute Supreme Court of Pakist UK United Nations US and British embassy US Army US Circuit Court of Appe	Brussels Bundesland Karnten Bundesland Salzburg Bundesland Steiermark Bundesland Wien California City of London Douglas Hollywood Innsbruck Kingdom of Spain Kingdom of Sweden Kingdom of the Netherlar Linz London Los Angeles Poland Portuguese Republic Rome Sacramento Split U K United States Vienna
Recent Items <input type="checkbox"/> Republic of Austria				

Co-occurrence...execution

CORE Search

[Clear](#) | [Options](#)

Document Keyword Filter	Key Phrases	People	Organizations	Locations
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>
Matching documents: 5 <input type="button" value="Documents"/> <input type="button" value="Timelines"/>	20 matching entities	21 matching entities	8 matching entities	15 matching entities
Selected Items Filter <input type="checkbox"/> Arnold Schwarzenegger <input type="checkbox"/> Graz <input type="checkbox"/> Republic of Austria	case city death Execution governor honour injection innocence name news agency picture republic reputation request resident ring son sport stadium sun victory	Archbishop Rowan Willie Benita Ferrero-Waldner Christopher D. Hughes Denzel Washington Dieter Hardt-Stremayr Donald Beardslee Joerg Haider Karl Kling Matthew Cooper Monika Ficszko Mozart Peggy Ryen Pilz Quentin State Prison Sean Penn Siegfried Nagl Stanley Thomas Rajakovics Waltraud Klasnic Wolfgang Schuessel	AFP Austrian TV Court of Appeals Green Party Reuters Group PLC Supreme Court of Pakist US and British embassy US Circuit Court of Appe	Bundesland Karnten Bundesland Steiermark Bundesland Wien California Commonwealth of Austr Douglas Europe Hollywood Liebenau Stadium Los Angeles Sacramento Schwarzenegger Stadiu United States
Recent Items <input type="checkbox"/> Republic of Austria				

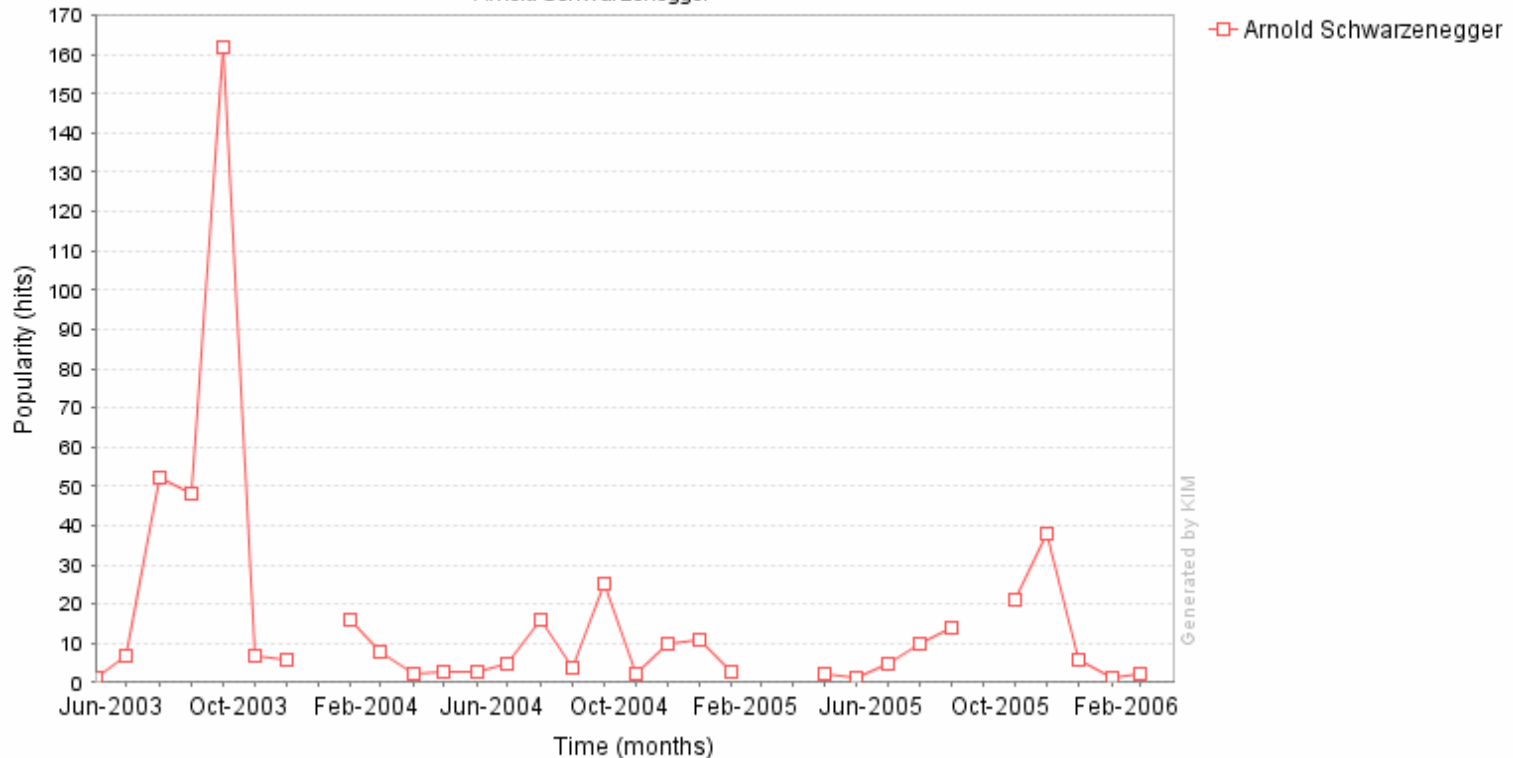
Input Form

Arnold's Popularity

Timelines Result

Timelines for entities :

Arnold Schwarzenegger



Document Filter: 167 docs, containing Keywords : (none) Entities : Arnold Schwarzenegger

Time Period: 01/01/2003 to 31/03/2006 Granularity: Month

Documents Forming the Peak

Document Query Result

Date	Title
07/10/2003 21:46	Q&A: California recall vote
06/10/2003 08:06	Profile: Cruz Bustamante
29/10/2003 04:31	Forecast brings hope for stricken state
29/10/2003 20:17	Firefighter dies in California blaze
09/10/2003 06:40	Arnie steals Ugandan thunder
02/10/2003 20:15	Sex scandal draws Arnie apology
03/10/2003 09:47	Gropes not arms interest US
07/10/2003 21:47	Californians make up their minds
08/10/2003 03:43	Schwarzenegger win dominates US media
09/10/2003 04:14	Maria Shriver: Arnie's secret weapon
09/10/2003 04:20	World press digests Arnie's latest role
07/10/2003 21:33	US diary: California recall
08/10/2003 04:08	Arnie wins California election
08/10/2003 02:02	Arnie win 'sign of US discontent'
10/10/2003 02:49	Arnie names transition team

1-15 of 32 Documents per page:

Available Query Options

Timelines

Document Distribution



Take home messages from this talk

- RDF based information repositories are a new breed of database which combine large-scale storage with machine reasoning
- OWLIM is one such new RDF repository based on the Sesame Open Source RDF database, and adding commercial services to it
- We can increasingly, use information retrieval techniques to enhance text documents with machine-understandable descriptions
- Future information systems will combine models and rules, text-based retrieval, data, transactions and machine reasoning
- Using XML and RDF for data encoding is safe for interoperation
- Ontologies, schemas, catalogues, metadata and “pure” data can be confusing. think instead:
 - Does a statement S concern an individual?
(then it is data)
 - Does statement S hold for a collection or set of individuals?
(then it is a schema)
 - Conclusion: one observer’s schema can be another observer’s data!
- Natural language text is already a form of knowledge representation!

Thank You!

<http://www.ontotext.com/>