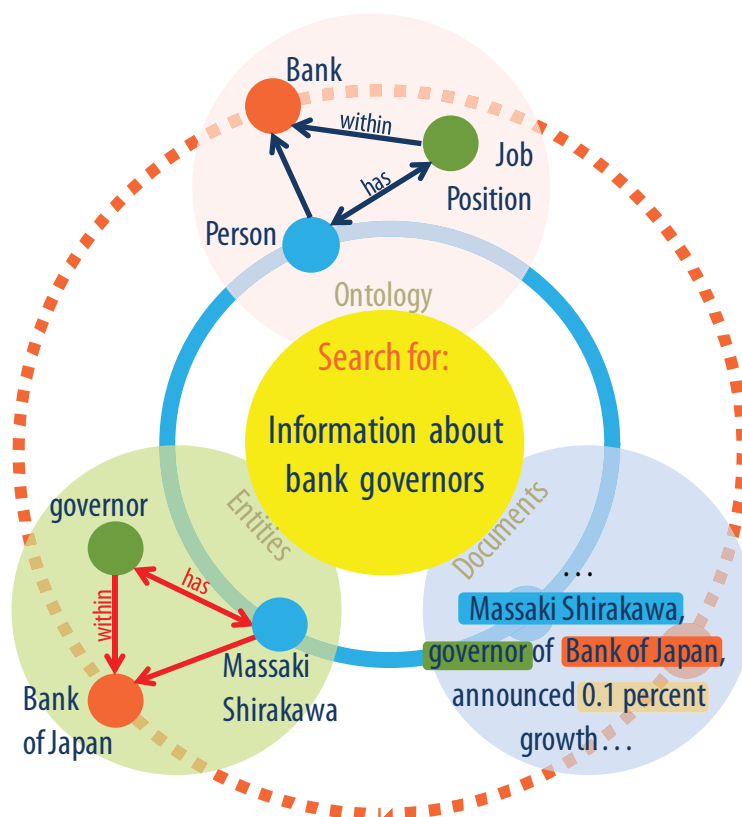# SEMANTIC ANNOTATION and SEARCH

**kim**

## FINDING WHAT YOU NEED

*Large amounts of documents and data >> multiple search paradigms >> rich results*

**KIM EMPOWERS YOU TO LINK, SEARCH, AND NAVIGATE** your information space to find content, data, and facts.

**KIM CREATES SEMANTIC LINKS** between your documents, data, domain models (ontologies), and linked data, finding mentions of entities, relationships, and facts in texts.

As a platform, **KIM IS HIGHLY CUSTOMIZABLE** for different data, domains, search and navigation needs. It scales up as your needs grow and can be smoothly integrated in your existing software environment.
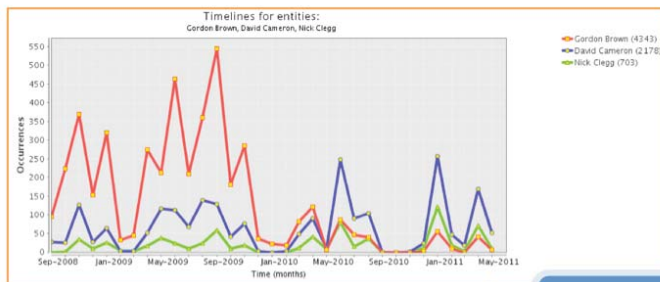


**ontotext**

# SEMANTIC ANNOTATION and SEARCH

## HOW TO DO IT
*Select entities* >> *find relations* >> *get rich results*



**Timelines for entities:**
Gordon Brown, David Cameron, Nick Clegg

- Gordon Brown (4343)
- David Cameron (2178)
- Nick Clegg (703)

### Space

| | |
|---|---|
| Office **Space** | [General Term] |
| Russian **Space** Agency | [Organization] |
| Http://**space**flight.nasa.gov | [Web Page] |
| **Space** | [Organization] |
| National Aeronautics and **Space** Agency | [Government] |
| Goddard **Space** Flight Center | [Government] |
| Johnson **Space** Center | [Organization] |

**LOOK FOR ENTITIES**

**ANALYZE POPULARITY TRENDS IN A DATA SUBSET**

**GET SUGGESTIONS**

### Selected Items
- Gordon Brown
- David Cameron
- Nick Clegg
- Republic of Iraq
- United Nations

**EXPLORE ENTITY DESCRIPTIONS OR DOCUMENTS MENTIONING THEM**

**SELECT ENTITIES**

### Title
Yahoo Reports a Profit, but Revenue F
**SAN FRANCISCO** (AP) Yahoo annou
Yahoo names tech veteran Carol Bartz
Yahoo names tech veteran Carol Bartz
Google loses board member over links
Google loses board member over links
...organisations. But although his depa
Appeals court orders new trial in Broca
Appeals court orders new trial in Broca
against ...
Apple bucks recession with some of be
...best financial results in its history, th
...last week that the Seattle technolog
Engineers sentenced to 1 year for espi
...were sentenced to a year in prison f
...case against Ye and Zhong stretche

**LOOK FOR ENTITIES MENTIONED IN THE SAME TEXT**

**GET DOCUMENTS**

- California Amplifier, Inc.
- Call-Net Enterprises Inc.
- D&E Communications, Inc.
- Eagle Broadband Inc.
- EarthLink, Inc.

**GET A LIST OF RELATED ENTITIES (OR DOCUMENTS)**

**EXPLORE ENTITY DESCRIPTIONS**

**EPLOY RELATIONS**

**FIND RELATED ENTITIES**

### San Francisco is a City, Trusted[tip!]
has Alias **San Francisco**
has Main Alias **San Francisco**
Part of **United States**, **California**, **North America**
Located in **United States**, **California**, **North America**
Subregion of **United States**, **California**, **North America**
rank "4583"

**Related Entities**
The Fed   Located in   San Francisco
the House   Located in   San Francisco
Federal Reserve Bank   Located in   San Francisco
Government   Located in   San Francisco
FDIC   Located in   San Francisco
Securities and Exchange Commission   Located in   San Francisco

| Company | |
|---|---|
| active in Industry Sector | Telecom |

**Telecom**munications  Industry Sector

## ontotext

http://www.ontotext.com/kim

# WHAT WE CAN DO FOR YOU

We can analyze your needs and fully tailor KIM to suit you by:

ADAPTING THE EXTRACTION METHODS

DESIGNING AND INTEGRATING RELEVANT ONTOLOGIES

SEMANTICALLY INTEGRATING YOUR VARIOUS INFORMATION AND DATA SOURCES

TRANSFERRING KNOWLEDGE TO YOUR TECHNICAL TEAM

# APPLICATIONS AREAS

ENTERPRISE SEARCH AND SEMANTIC CONTENT MANAGEMENT: linking multiple sources of structured data and texts to provide an integrated view over your data

HEALTHCARE, PHARMACEUTICAL AND CHEMICAL INDUSTRIES: annotation, search and semantic data integration in clinical studies, patient records, scientific articles, and patents

NEWS ENRICHMENT for major media companies

LINKED DATA BASED ANNOTATION AND SEARCH: for public and corporate documents

ENTITY PROFILING: based on profile feeds like World Check and Dow Jones, and extending profiles automatically with what appears about people, organizations and their relations on the Web, to be applied for asset recovery with several governments' financial intelligence units

REGULATIONS AUTHORING, MANAGEMENT, AND SEARCH: ontology-driven searching capabilities that will vastly improve access to key regulatory information for US government users
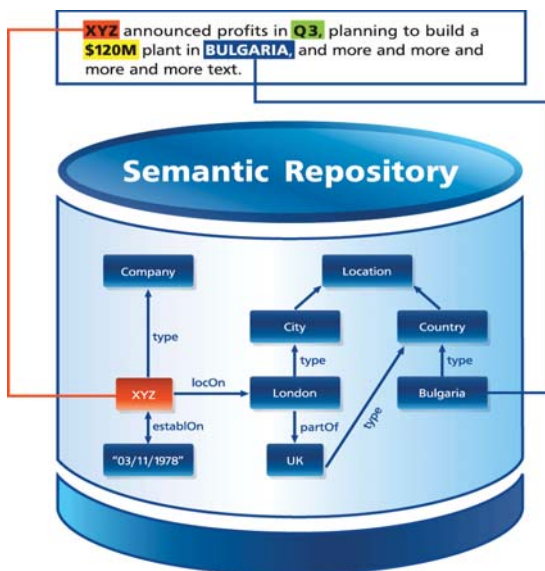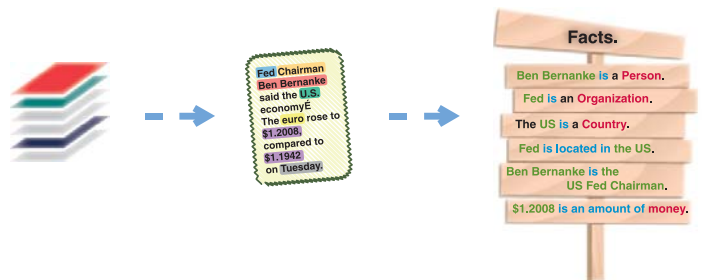
## TEXT ANALYSIS

*Automatic extraction of facts >> linking content and structured knowledge >> advanced filter and search*

### INFORMATION EXTRACTION

To link your different types of data, KIM employs text mining, finding entity and relation mentions in the analyzed content.

It combines symbolic and statistical methods and can be adapted to any domain.



Fed Chairman Ben Bernanke said the U.S. economy. The euro rose to $1.2008, compared to $1.1942 on Tuesday.

Facts.
Ben Bernanke is a Person.
Fed is an Organization.
The US is a Country.
Fed is located in the US.
Ben Bernanke is the US Fed Chairman.
$1.2008 is an amount of money.



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

### SEMANTIC ANNOTATION

The extracted metadata is further linked to relevant conceptual models to create a semantic index over content, facts and background knowledge in the system.

Semantic annotation goes beyond the mere categorization of entity mentions. It determines their identity through disambiguation and matching to knowledge base instances.

# ARCHITECTURE

*Modular >> scalable >> highly customizable*

KIM is a highly adaptable and modular platform for linking and navigating data, content, and knowledge. It can be configured to use all or some of its components to suit different needs. Computationally intense components as concept extraction and semantic database can be clustered to reach the performance you need.
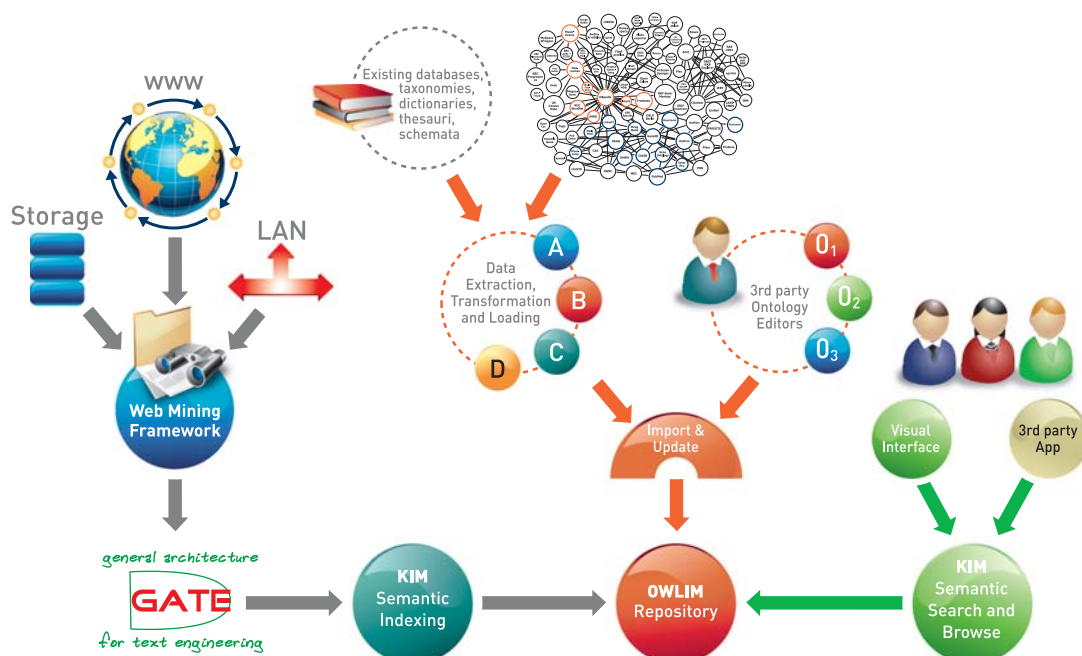
ACQUIRES CONTENT through focused Web crawling or import from file storage

ORGANIZES AND UPLOADS KNOWLEDGE through mapping of ontologies, taxonomies, Linked Data

EXTRACTS FACTS from content

CREATES SEMANTIC INDEX linking content and knowledge

PROVIDES MULTI-PARADIGM SEMANTIC SEARCH over your data, documents, and facts



Each of the major platform components can be customized and used independently in a 3rd party system. KIM is based on GATE (text analysis components) and OWLIM (semantic index).

# UK GOVERNMENT WEB ARCHIVE GOES SEMANTIC · kim

SHOWCASE

*Multi-paradigm indexing* >> *advanced filter and search*

Ontotext together with the GATE team from the University of Sheffield and System Simulation Ltd. were granted a project to create a semantic knowledge base (SKB) to improve the search and navigation in the UK Government Web Archive. SKB interlinks the UK Government Web Archive with DATA.UK.GOV and other LOD datasets to deliver semantic search at ultimate scale: 150 million documents annotated against 5 billion triples.

KIM was applied to the National Archive project to provide:

**SOFTWARE FRAMEWORK FOR MULTI-PARADIGM INDEXING:** accessing and linking together multiple information sources, such as textual content of documents, the semantic metadata associated with the documents and data in knowledge bases, such as the linked data cloud

**ADVANCED FILTER AND SEARCH:** pattern search, faceted search, semantic suggestions and timelines; RDF-level exploration and SPARQL querying

*"The Semantic KB implementation delivered by Ontotext solves two problems: the collection contains a large amount of near-duplicate material which makes conventional search tools not very effective; another major issue is how users continue to find topical information when government structures change over time"*

Amanda Spencer, Head of Web Continuity

ontotext

# BBC SPORT – DYNAMIC SEMANTIC PUBLISHING

## SHOWCASE

*Automatic concept extraction* >> *linking assets and data* >> *concept relevance ranking*

BBC's 2010 World Cup web site was the first project running on Dynamic Semantic Publishing (DSP) architecture. DSP provided an improved navigation, content re-use and re-purposing, and support of semantic advertisement.

OWLIM serves as an RDF store in the middle of the architecture facilitating metadata-based management and retrieval of content. Later on, Ontotext also implemented another important part of the DSP architecture - the Concept Extraction Service.

Based on GATE, the Concept Extraction Service provides "advanced methods for analysing content and deciding how to tag this content with precise  metadata linked to uniquely identified concepts (a concept usually being a person, place or thing)".

The main benefits of the Concept Extraction Service are:

AUTOMATIC LINKING of assets and concepts.

DYNAMIC EXTRACTION ACCURACY IMPROVEMENT based on journalist curation.

MULTI-COMPONENT RELEVANCE RANKING of extracted concepts.

*"A high-performance dynamic semantic publishing framework. This framework facilitates the publication of automated metadata-driven web pages that are light-touch, requiring minimal journalistic management, as they automatically aggregate and render links to relevant stories."*

*"The Concept Extraction Service developed by Ontotext becomes the main content enrichment facility for BBC SPORT"*

Jem Rayfield, Senior Technical Architect

http://www.ontotext.com/kim

**KIM PLATFORM**

## CONTACT INFORMATION

*If you are interested, you can subscribe to:*

KIM-info@ontotext.com

FOR GENERAL PURPOSE QUESTIONS, comments, feedback, etc.

interested-in-KIM@ontotext.com

FOR GENERAL INFORMATION about known issues, major updates, news, and etc.
Please note that you cannot reply directly to this mailing list address.

KIM-discussion@ontotext.com

FOR TECHNICAL QUESTIONS, issues, requests, and comments, and if you want to join the growing KIM Platform community.

**Ontotext USA Inc.**
2490 Black Rock Turnpike #331
Fairfield
Connecticut 06825-2400, USA

Contact: Mr. Matthew Petrillo
Tel. +1 (718) 785 9692
Toll-free for N. America:
+1 (866) 972 6686

**Ontotext AD**
135 Tzarigradsko Chaussee, Floor 3
1784 Sofia, Bulgaria

Contact: Ms. Kamelia Atanasova
Tel:  (+359 2) 974 61 60
Fax:  (+359 2) 975 32 26

**info@ontotext.com**
**www.ontotext.com**

http://www.ontotext.com/kim