Outline

credit:
Maciej Janik
based on material of:
Tim Finin and Mathieu d'Aquin

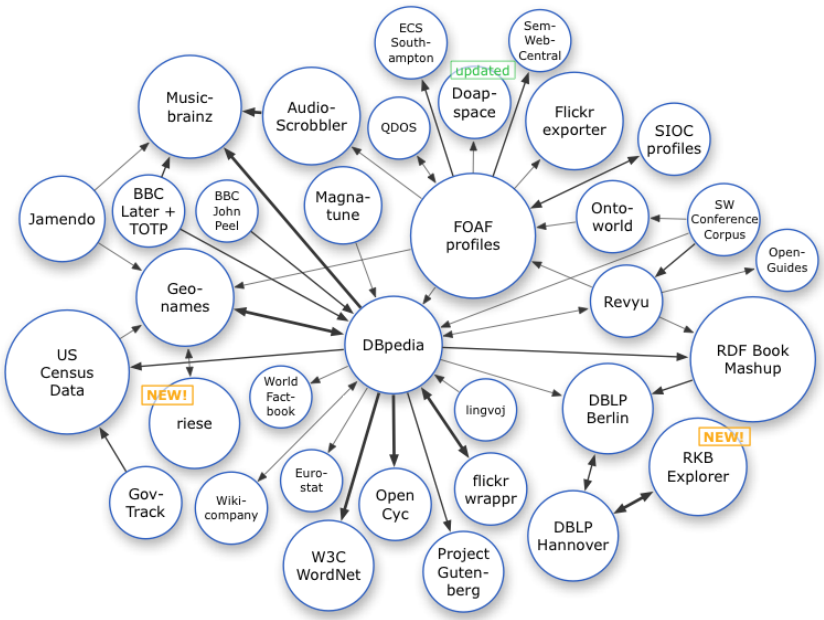
Outline

Types of semantic search engines

- Semantically-enhanced Search
  - Yahoo! SearchMonkey, Google squared (2009 - 2011)
- NLP-based Search
  - MetaWeb Freebase, Powerset,...
- Semantic-NLP-based Search
  - hakia, Cognition, ...
- Computational-NLP-based Search
  - True Knowledge, Wolfram Alpha, ...
- Semantic Web Search (search for ontologies)
  - Swoogle, Sindice, SWSE, Falcon-S, Watson, Shoe, ...

Semantic data vs. application needs

Data:



Applications:

- Discover how entities are related
- Browse ontologies
- Answer NLP questions
- Query multiple (distributed) knowledge bases
- Find and rank (relevant) ontologies

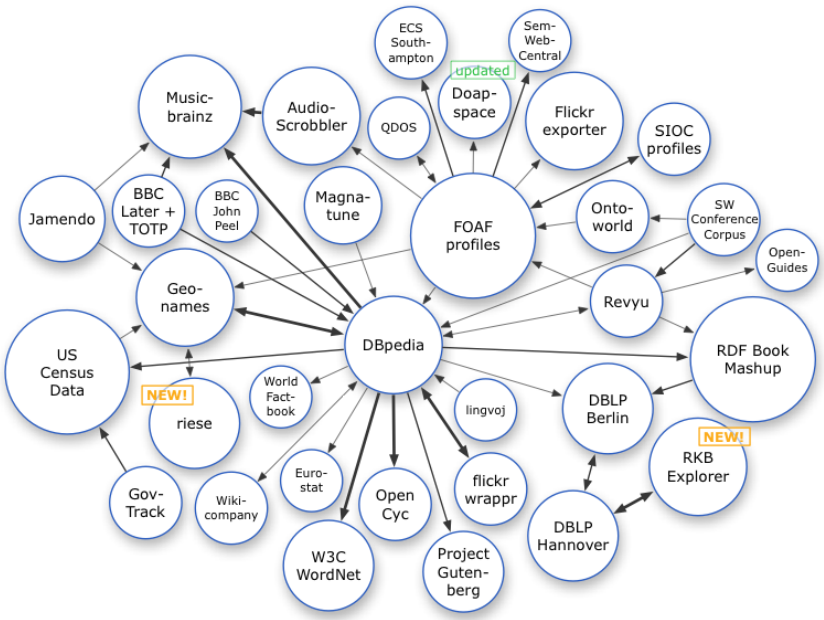
→ Gateway to semantic data: dynamically

- retrieving
- exploiting
- combining

relevant semantic resources

Semantic data vs. application needs

Data:



Applications:

- Discover how entities are related
- Browse ontologies
- Answer NLP questions
- Query multiple (distributed) knowledge bases
- Find and rank (relevant) ontologies

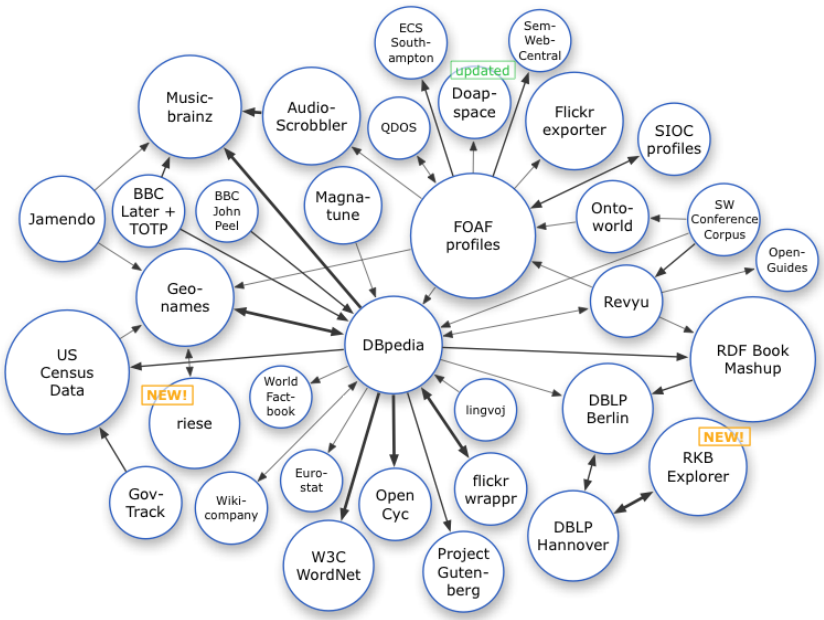
→ Gateway to semantic data: dynamically

- retrieving
- exploiting
- combining

relevant semantic resources

Semantic data vs. application needs

Data:



Applications:

- Discover how entities are related
- Browse ontologies
- Answer NLP questions
- Query multiple (distributed) knowledge bases
- Find and rank (relevant) ontologies

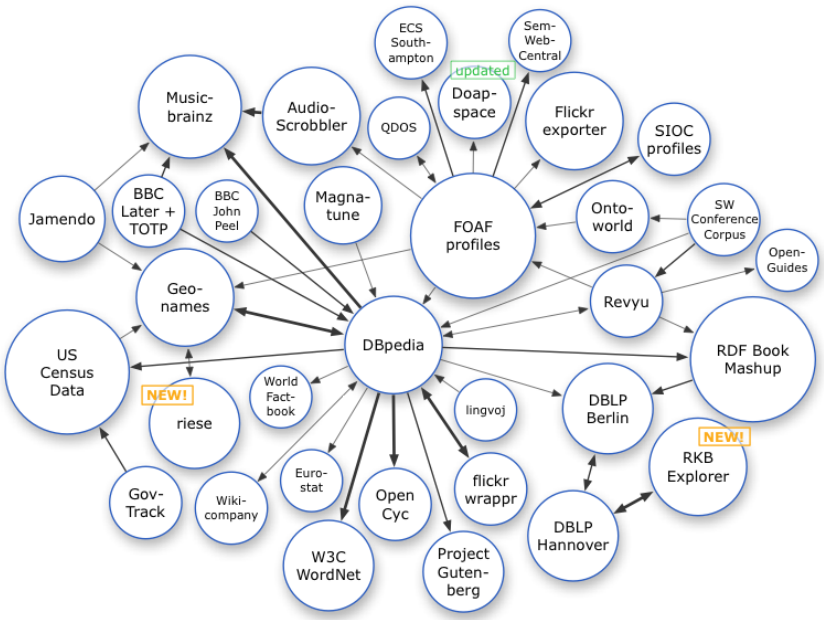
→ Gateway to semantic data: dynamically

- retrieving
- exploiting
- combining

relevant semantic resources

Semantic data vs. application needs

Data:



Applications:

- Discover how entities are related
- Browse ontologies
- Answer NLP questions
- Query multiple (distributed) knowledge bases
- Find and rank (relevant) ontologies

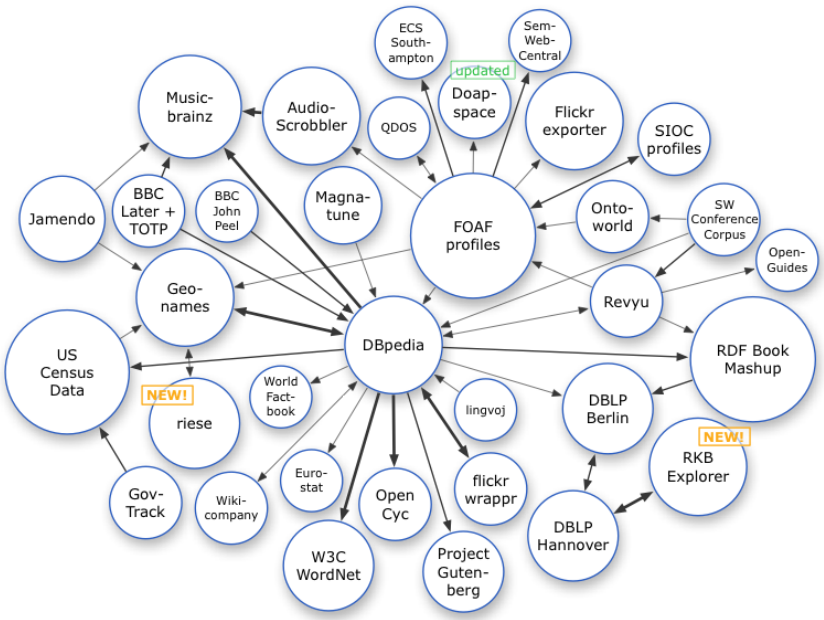
→ Gateway to semantic data: dynamically

- retrieving
- exploiting
- combining

relevant semantic resources

Semantic data vs. application needs

Data:



Applications:

- Discover how entities are related
- Browse ontologies
- Answer NLP questions
- Query multiple (distributed) knowledge bases
- Find and rank (relevant) ontologies

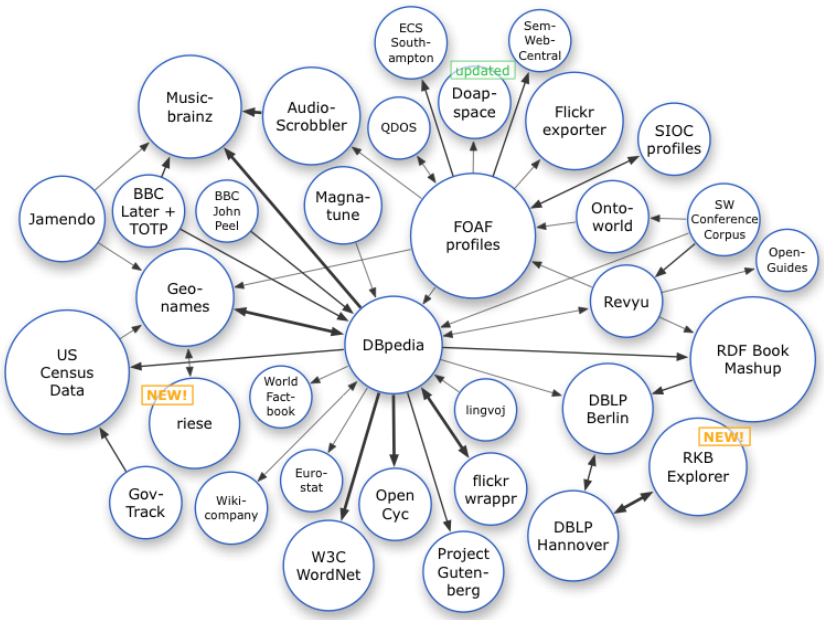
→ Gateway to semantic data: dynamically

- retrieving
- exploiting
- combining

relevant semantic resources

Semantic data vs. application needs

Data:



Applications:

- Discover how entities are related
- Browse ontologies
- Answer NLP questions
- Query multiple (distributed) knowledge bases
- Find and rank (relevant) ontologies

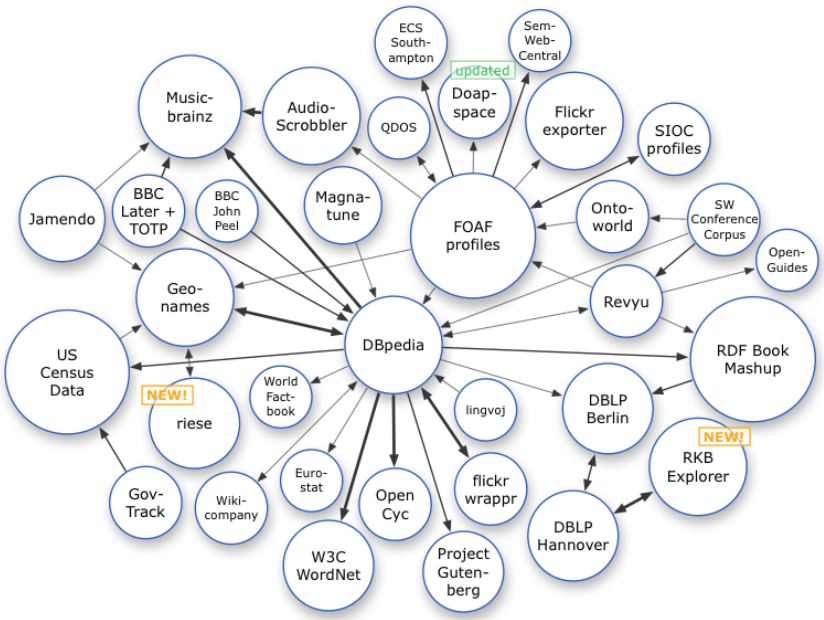
→ Gateway to semantic data: dynamically

- retrieving
- exploiting
- combining

relevant semantic resources

Semantic data vs. application needs

Data:



Applications:

- Discover how entities are related
- Browse ontologies
- Answer NLP questions
- Query multiple (distributed) knowledge bases
- Find and rank (relevant) ontologies

→ Gateway to semantic data: dynamically

- retrieving
- exploiting
- combining

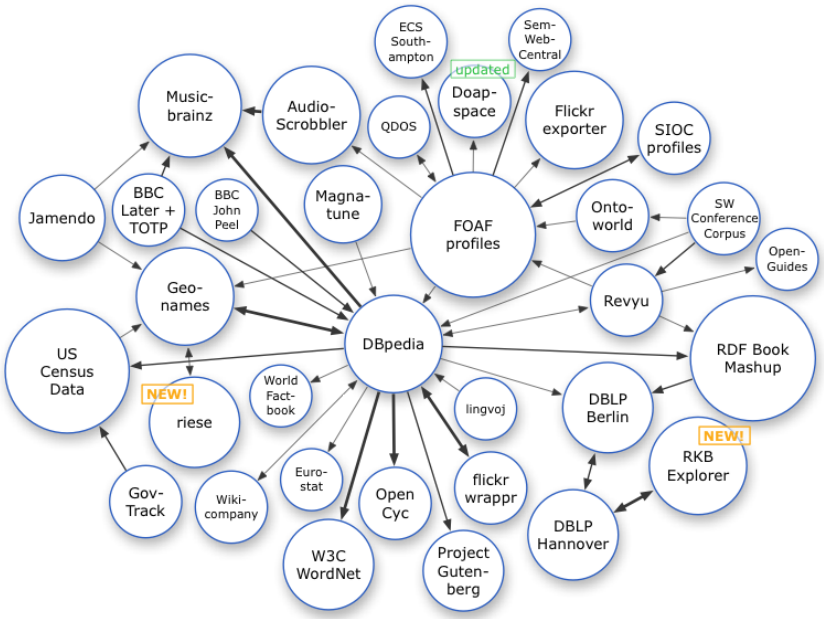
relevant semantic resources

Semantic Search? Finding Ontologies?

- For reuse
  - To build upon what exists
  - To adopt what is used in practice
  - Not to re-invent the wheel
  - Because it is simpler than building from scratch
- For applications
  - Because semantic applications need knowledge
  - Because knowledge is hard to acquire
  - Because some scenarios require to gather this knowledge at run-time
  - Because in some scenarios, the more there is, the better

Swoogle – the first semantic Web gateway

Data:



Applications:

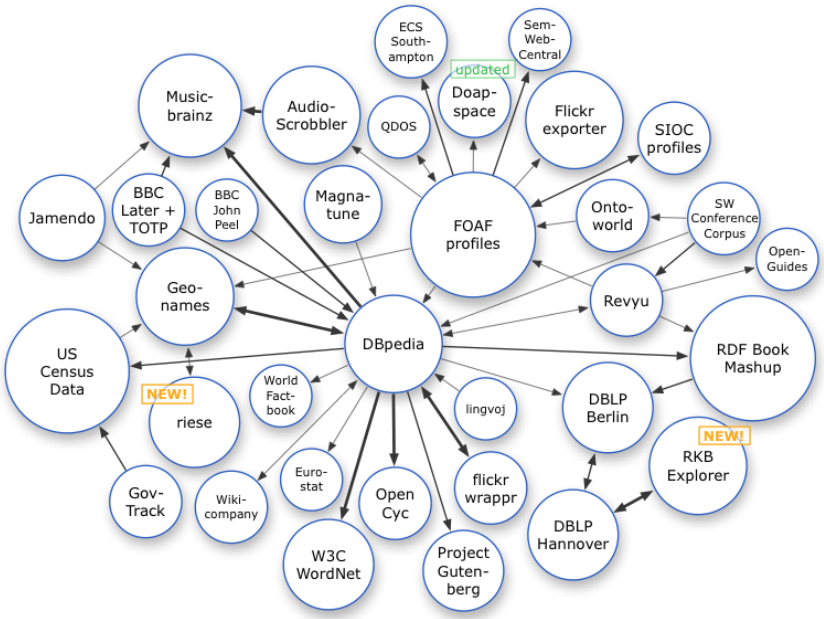
- Discover how entities are related
- Browse ontologies
- Answer NLP questions
- Query multiple (distributed) knowledge bases
- Find and rank (relevant) ontologies

→ Gateway to semantic data:

Swoogle
semantic web search

Swoogle – the first semantic Web gateway

Data:



Applications:

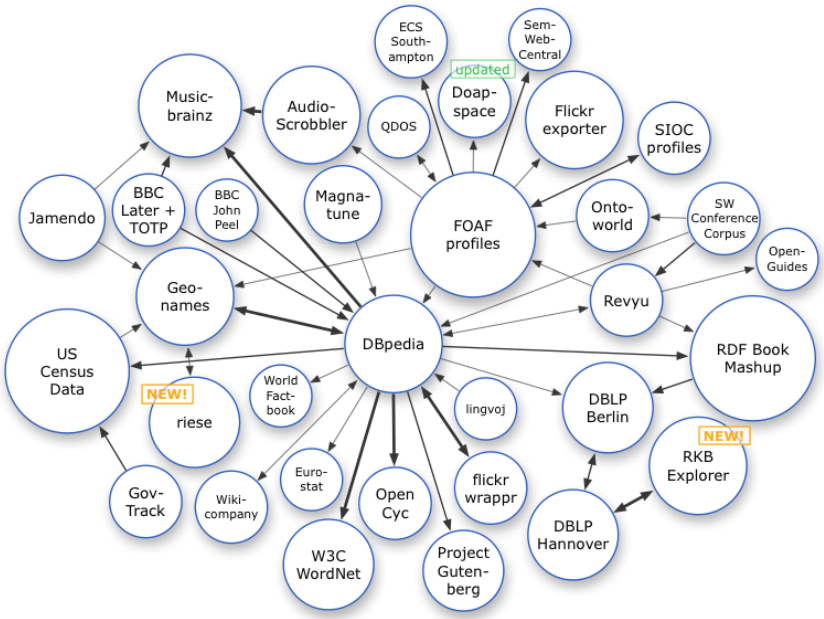
- Discover how entities are related
- Browse ontologies
- Answer NLP questions
- Query multiple (distributed) knowledge bases
- Find and rank (relevant) ontologies

→ Gateway to semantic data:

Swoogle
semantic web search

Swoogle – the first semantic Web gateway

Data:



Applications:

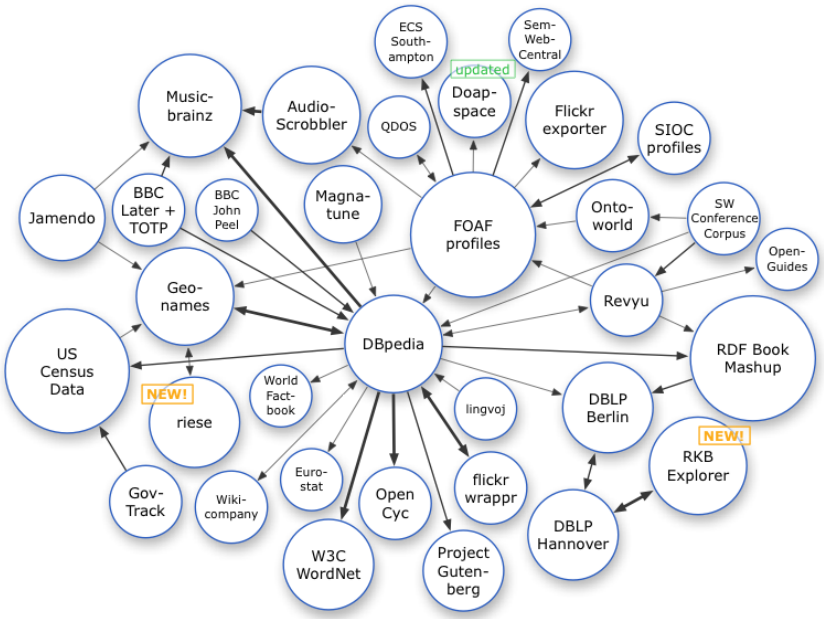
- Discover how entities are related
- Browse ontologies
- Answer NLP questions
- Query multiple (distributed) knowledge bases
- Find and rank (relevant) ontologies

→ Gateway to semantic data:

Swoogle
semantic web search

Swoogle – the first semantic Web gateway

Data:



Applications:

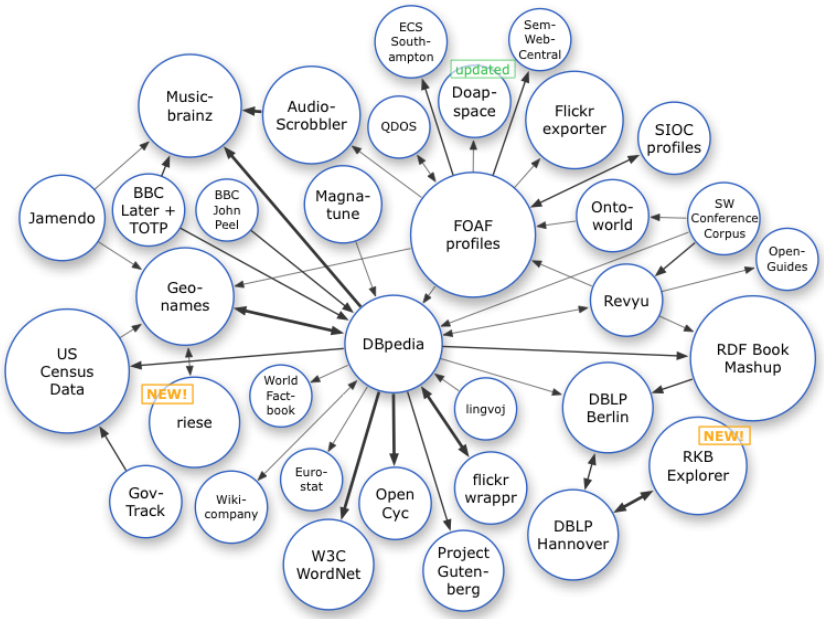
- Discover how entities are related
- Browse ontologies
- Answer NLP questions
- Query multiple (distributed) knowledge bases
- Find and rank (relevant) ontologies

→ Gateway to semantic data:

Swoogle
semantic web search

Swoogle – the first semantic Web gateway

Data:



Applications:

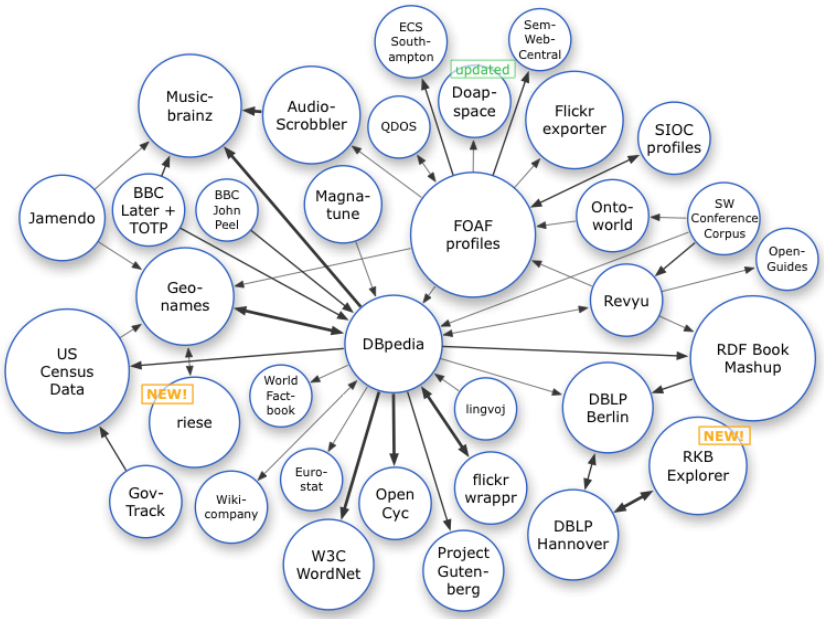
- Discover how entities are related
- Browse ontologies
- Answer NLP questions
- Query multiple (distributed) knowledge bases
- Find and rank (relevant) ontologies

→ Gateway to semantic data:

Swoogle
semantic web search

Swoogle – the first semantic Web gateway

Data:



Applications:

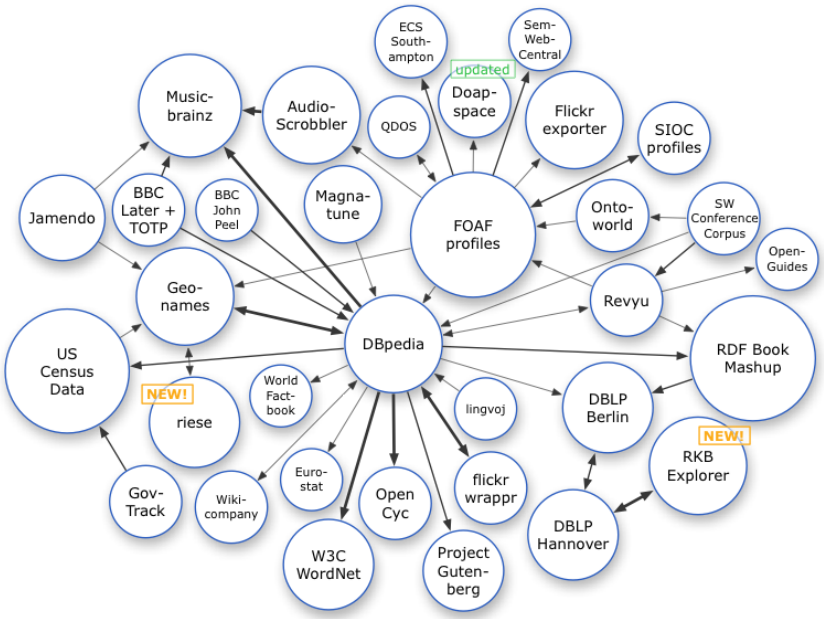
- Discover how entities are related
- Browse ontologies
- Answer NLP questions
- Query multiple (distributed) knowledge bases
- Find and rank (relevant) ontologies

→ Gateway to semantic data:

Swoogle
semantic web search

Swoogle – the first semantic Web gateway

Data:



Applications:

- Discover how entities are related
- Browse ontologies
- Answer NLP questions
- Query multiple (distributed) knowledge bases
- Find and rank (relevant) ontologies

→ Gateway to semantic data:

Swoogle
semantic web search

Swoogle

- Created in 2004
- Crawls and discovers documents in RDF, OWL
- Indexing and retrieval system
- Search for Semantic Web Documents (SWD)
  - Ontologies
  - Instance data

Swoogle — search and metadata for the semantic web

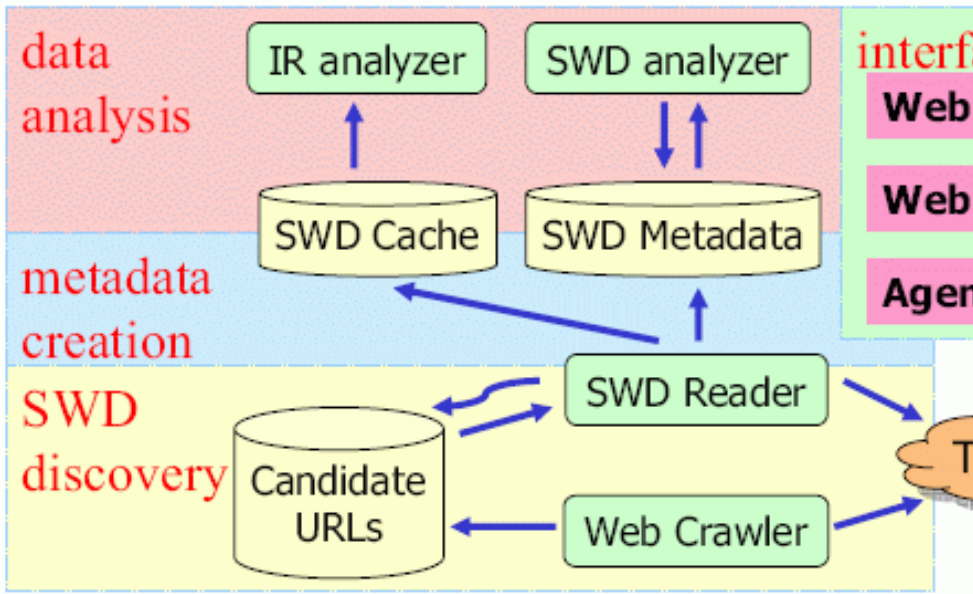
- Created in 2004
- Crawls and discovers documents in RDF, OWL
- Indexing and retrieval system
- Search for Semantic Web Documents (SWD)
  - Ontologies
  - Instance data

How Swoogle can be used?

- Find ontologies
  - Containing keywords, terms, concepts
  - Similar to 'http://myontology.org/...'
  - Used to describe document X (directly or indirectly)
- Find semantic Web documents
  - Containing keywords, terms ...
  - Used or created by specific institution
- Browse
  - Ontologies using specific topic hierarchy
  - Ontology metadata
  - Entities and navigate between them

Swoogle architecture

- Uses Google APIs to discover URIs
- Crawls using these URIs as seeds
- Allows users to submit URIs
- Offers multiple interfaces
- Generates Metadata
  - Used for ranking
  - Basic RDF statistics and ontology annotation.
  - RDF Statistics (determine SWD or SWO)
  - Ontology annotation



Limitations of Swoogle

- No quality control mechanisms
  - Many ontologies are duplicated
  - No quality information provided
- Limited Query/Search mechanisms
- No support for **relations between ontologies**
  - Duplication, incompatibility (contradiction), modularization, versioning, etc.

Computational NLP Search: Wolfram Alpha

→ **computational** knowledge engine

- often referred to as "semantic search engine"
- the purpose is **not** to find appropriate Web pages as result
  - Instead: *Compute* a a result (an answer based on the content of Web sites) to a query.

  → achieved by composing results

Semantic NLP-based Search: hakia

- Motivation: conventional search engines are no longer appropriate to provide high quality results.

$\rightarrow$ expect of matching occurrences of words (terms), the idea is to take the meaning of terms into account.

- Benefits of "semantic search":
  - Enables semantic *understanding* of users
  - Accuracy of results
  - Focus (of the result) is on *answering a question* rather than on *finding relevant Web pages* (cf. like in Wolfram Alpha)

Structured knowledge base for NLP-based Search: Freebase

- Freebase: collaborative knowledge base

→ online collection of structured data harvested from many sources, including individual 'wiki' contributions
  - data is harvested from sources like Wikipedia, MusicBrainz.

- the data model is a graph

- Freebase follows the principle of *folksonomies*, in which people can just add new terms (topics) like tags.
- Different to the Wiki-like approach ...
  - users can create their own types
  - but there is some control mechanism by Freebase

Semantically-enhanced Search

- Idea: Use structured data (in search engines) to make results more useful.
- Examples: Yahoo! SearchMonkey
- Typically, additional structure and information is provided, e.g., in terms of *microformats*
  → see *schema.org*