Introduction to Information Retrieval http://informationretrieval.org

Semantic Search

### Hinrich Schütze & Wiltrud Kessler

Institute for Natural Language Processing, Universität Stuttgart

2013-01-22



- 1 The Semantic Web
- 2 Semantic Search
- Inhancing Keyword Search
- 4 Searching for Concepts and Relations
- **5** Beyond Document Retrieval



The Semantic Web	Keywords	Concepts	
Outline			

- 1 The Semantic Web
- 2 Semantic Search
- Inhancing Keyword Search
- 4 Searching for Concepts and Relations
- Beyond Document Retrieval

### 6 Summary

The Semantic Web Vision

ywords

epts Entities

Summar



The vision [Tim Berners-Lee et al., The Semantic Web, Scientific American, May 2001]

"The Semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation."

- The Semantic Web is a global database with semantic information interpretable by humans and machines enabling the easy exchange of data.
- The Semantic Web is driven by Tim Berners-Lee and the World Wide Web Consortium (W3C), an international standardization body for the Web.

Semantic Search

Keywords

Concepts

Summary

### The Semantic Web Cake [W3C, Tim Berners-Lee]



Keywords

Concepts Er

Summary

### The Layers of the Cake

- Data is identifiable and uniformly encoded
  - $\rightarrow$  URI and Unicode
- Data is described in a uniform syntax
  - $\rightarrow$  XML and XMLSchema
- $\bullet$  Simple relations between data items can be expressed  $\rightarrow$  RDF and RDFS
- Data semantics is formally defined
  - $\rightarrow$  Ontologies and OWL
- Information can be queried, new information can be inferred  $\rightarrow$  SPARQL and Logic
- Information sources can be verified and are trusted
  - $\rightarrow$  Crypto and Trust
- Applications use the provided information and mechanisms
  - $\rightarrow$  Software Agents

epts Entities

Summary

### Resource Description Framework (RDF)

- RDF is a data-model.
- Its basic building block are object-attribute-value triples, called **statements**.
- The RDF statements form a (directed) graph.
- The nodes are **resources** are the things we talk about. Examples: Books, lectures, people, places, ...
- The edges are **properties** that describe a relation between two resources.

Examples: written by, age, title, ...







- RDF allows us to talk about relations between individuals.
- RDFS and OWL provide a vocabulary to talk about classes and their relations → ontologies.
- Examples:

Instance of An individual is member of some class. Example: John Smith is a Person.

Hierarchy Every instance of a subclass is also an instance of the superclass.

Example: Customer is a Person.

Property Restrictions To be a member of a class, an instance has to fulfill some restrictions. Example: A vegetarian pizza may not contain a

Example: A vegetarian pizza may not contain a meaty topping.

The Semantic Web Semantic Search Keywords Concepts Entities Summary

### The Pizza Ontology





- The terms of RDFS/OWL have a fixed semantic interpretation defined in first-order predicate logics.
- It is not up to the application to interpret these terms
   → RDFS makes semantic information machine-accessible.
- We can use reasoning methods from predicate logics to infer new knowledge from the given facts.
   Example: If John Smith is a customer, and customer is a person, John Smith is a person.



- 1 The Semantic Web
- 2 Semantic Search
- Inhancing Keyword Search
- 4 Searching for Concepts and Relations
- Beyond Document Retrieval
- 6 Summary

### How can Semantic Web Technologies help?

- Traditional information retrieval is keyword-based.
- No interpretation of the "meaning" of the information.
- Problems of this basic approach:
  - Polysemy (jaguar the cat vs. jaguar the car)
  - Synonyms (movie vs. film)
  - Missing information about subclass or part-of relation (watersport vs. diving, surfing, ...)
  - Relations between search terms ("books about recommender systems" vs. "systems that recommend books")
- This is where Semantic Web technologies can help!

### Semantic Search

### Working definition: Semantic Search [HHLV09]

Semantic Search is a process of information access, where one or several activities can be supported by a set of functionalities enabled by semantic technologies.

- Search engine functionalities: query construction, query processing, result presentation
- Semantic technologies: knowledge extraction, knowledge representation, reasoning.



- The Semantic Web
- 2 Semantic Search
- Inhancing Keyword Search
- 4 Searching for Concepts and Relations
- Beyond Document Retrieval

### 6 Summary





### Semantic Search on a Keyword Index

Semantic Search

- Entries in the dictionary are keywords (like in traditional information retrieval).
- Additionally, we have an ontology.
- Keywords map to elements in the ontology.

Keywords

• The ontology is used to disambiguate the query, e.g., to select the right word sense, and to expand the query.

Entities

• As a result, the mapped concepts and documents containing the keywords are returned.

The Semantic Web

oncepts E

### Example: CleverSearch (1)

Word Sense disambiguation using WordNet.

The word Java has 3 senses.

0	Reti	ırn all senses
OR		
۲	Reti	ırn only this sense:
	0	an island in Indonesia to the south of Borneo; one of the world's most densely populated regions
	۲	<b>coffee, java</b> a beverage consisting of an infusion of ground coffee beans; "he ordered a cup of coffee"
	0	a platform-independent object-oriented programming language
Sut	omit	

[http://wdok.cs.uni-magdeburg.de/clever-search/]

Keywords

Concepts

### Example: CleverSearch (2)

The query is expanded with synonyms and hyponyms (more specific terms) and a normal Google search with the expanded query is performed.

Google	Java" "coffee" OR "coffee substitute" OR "Irish coffee" OR "cafe au lait" OR '
Suche	Ungefahr 358.000.000 Ergebnisse (0.26 Sekunden)
	Alle Wörter ab "Turkish" wurden ignoriert, da Suchanfragen auf 32 Begriffe beschränkt sind.
Alles	Tipp: <u>Suchen Sie nur nach Ergebnissen auf <b>Deutsch</b>. Sie können Ihre Sprache in den <u>Einstellungen</u> festlegen.</u>
Bilder	
Maps	Java coffee - Wikipedia, the free encyclopedia en.wikipedia.org/wiki/Java_coffee - Diese Seite übersetzen
Videos	Java coffee is a coffee produced on the island of Java. In the United States the term "Java" by itself is, in general, slang for coffee. The Indonesian phrase Kopi
News	
Shopping	Benutzer:Java-Contee – Wikipedia
Mehr	ucmingeona og min Denazo-Fonza Son Control en anter Son
Das Web Seiten auf Deutsch Übersetzte Seiten	Cafe au Lait Java News and Resources www.cafeaulait.org/-Diese Seite übersetzen 15 Nor 2010 – Cafe au Lait is the preeminent independent source of <b>Java</b> information on the net. Unike many other Java sites, Cafe au Lait is neither Brewing Java: A Tutorial - Java Course Notes - Java Book List - Java FAQ

### Possible Ways for Query Expansion

Keywords

Semantic Search

• The query can be expanded with words/concepts from a thesaurus (e.g., WordNet):

Entities

• Synonyms.

The Semantic Web

- More specific concepts.
- More general concepts.
- Related concepts, e.g., antonyms, meronyms, ...
- The query can be expanded with words/concepts from a domain ontology:
  - All of the above.
  - Concepts that are connected to the query concepts by a domain property, e.g., add "Milk" to a query about "Coffee" because it is a possible ingredient, add the author of a book to the search for the book.
- The description of concepts can be used to expand the query.



- The Semantic Web
- 2 Semantic Search
- Inhancing Keyword Search
- 4 Searching for Concepts and Relations
- 5 Beyond Document Retrieval

### 6 Summary

Entities

### A Document as Concepts and Relations



- A document doesn't contain keywords, it discusses concepts that are in a relation with the document.
- Concepts serve as as description of the documents for some known properties, e.g., type of document, author, ...

# The Semantic Web Semantic Search Keywords Concepts Entities Summary Several Keyword Indices

- Use one index for every relation (field) we are interested in.
- Possible relations are specified in an ontology.
- Relations may depend on the type of document (most application only support a specific class from a small domain, e.g., scientific documents, recipes).



### Searching for Concepts and Relations

Semantic Search

• The user decides what type of results (class in the ontology) he is looking for.

Concepts

Entities

- Properties of this class in the ontology can be used to narrow down results ("faceted search", each property is one facet).
- The possible values of the facet can be literals or other concepts, they can be restricted by the user.
- The ontology hierarchy relations can be used for inference (search for companies includes all subclasses/instances).
- As a result, documents from the class that match the property restrictions are returned.

The Semantic Web

eywords

Concepts

Summar

Entities

### Example: Yummly Recipe Search



[http://www.yummly.com/recipes]

The Semantic Web	Keywords	Concepts	Entities	
Outline				

- 1 The Semantic Web
- 2 Semantic Search
- Inhancing Keyword Search
- 4 Searching for Concepts and Relations
- **5** Beyond Document Retrieval

### 6 Summary



- If we have an ontology to describe the domain, why not use it to store the documents as well?
- The result is a (huge) RDF/OWL graph, we need repositories designed to deal with this kind of data.
- A semantic repository (or triple store) allows for storage, querying and management of RDF/OWL data (similar to database systems for relational data).
- Well-known implementations are e.g., Sesame (http://www.openrdf.org/) or Jena (http://jena.sourceforge.net/).

ncepts Entities

Summary

### Semantic Repository – Example



### How to Create a Semantic Repository

Several possible ways:

- Use semantically annotated data.
- Process documents with NLP techniques, e.g., Named Entity Recognition, Information Extraction, Relation Extraction, ...
- Import or access other existing ontologies / RDF documents, e.g., DBPedia (http://dbpedia.org), OpenCyc (http://sw.opencyc.org/), ...
- Create your own ontology / RDF documents.

.eywords

ncepts Entities

Summary

### Semantically Annotated Data

'''Nigeria''', officially the '''Federal Republic
of Nigeria''', is a [[Wikipedia:country|country]] in
[[located in::Africa|West Africa]] and the most populous
country on the [[Wikipedia:Africa|African continent]].
Nigeria shares land [[Wikipedia:border|border]]s with the
Republic of [[borders::Benin|Benin]] in the west, ...

- Markup to add semantic annotations.
- There is software to make the annotation easy, e.g., Semantic Media Wiki (http://semantic-mediawiki.org/).

- The user is not looking for documents, the user has some information need we try to answer by returning documents that may contain the answer.
- In a semantic repository, we have formalized knowledge about the relations between resources with RDF/OWL.
- We can directly try to identify the resource (entity) the user is looking for and return it with all associated information.
- This poses new challenges: How to recognize/map entities, how to rank them, ...

Semantic Search

Keywords

epts Entities

Summary

### Semantic Repository – Entity-centric



### Entity List – Example: SWSE (1)



[http://swse.deri.org/]

### Entity List – Example: SWSE (2)

#### George W. Bush

http://dbpedia.org/resource/George W. Busher

#### LABEL

- George W. Bush@de
- George W. Bush@en
- George W. Bush@es MORE...

#### PREFERRED LABEL

Bush, George W@en

#### COMMENT

- o, född 6 juli 1946 i New Haven, Connecticut, USA, är en amerikansk republikansk politiker som var USA:s 43:e president 2001–2009. Bush är son till George H.W. Bush som var 1989–1993 och 1981–1989. Före sin politiska karrär var Bush affärsman inom oljebranschen och delägare i basebollaget Texas Rangers. Bush valdes 1994 till Texas guvernör och omväldes 1988. 2000 blev han republikanermas presidentkandidat och vann över demokraten Al Gore i det jämna och omdiskuterade valet samma är.@sv
- o -- (英语:George Walker Bush, 岍, 1946年7月6日-),中国大陆译作布什,台湾译作布泰, 治澳译作布殊,朱、为美國第43任总统,布什在2001年1月20日詠寬,並且在2004年的選舉中擊脫民主黨參選人約喻,斯引該選連任,在擔任總統之前,希什於1995年至2000年間增任 第46任的德州州長,希什常族很早就開給敌入共和党說及美國政治,希什的父親是之前曾擔任第41任總統約方治 赫伯特·沃克·布什,他的弟弟貸布。布什也曾是佛羅里達州 約州唇,由於與父親同樣都員季國總統。及即父常讀嘉

#### IS IN SCHEME

o <u>nytd per</u>

Entities

#### SAMEAS

- 46571997880124065523
- bush george w per
- <u>George W. Bush</u> MORE...

#### TYPE

- Thing
- Agent
- military person

```
MORE...
```

#### DIFFERENTFROM

George H.W. Bush

#### SUBJECT

- <u>Category:George W. Bush</u>
- 20th-century American Episcopalians
- <u>Category:21st-</u> century presidents of the United ... MORE...

#### HOMEPAGE

- o index.html
- <u>www.georgewbushlibrary.com</u>
- www.georgewbushlibrary.gov

#### KNOWS

George H.W. Bush

```
PAGE
```

George W. Bush

USA TODAY - by Catalina Camia - 1 hour ago

Dallas Morning News (blog) - 1 day ago George W. Bush | Facebook

www.facebook.com/georgewbush

www.georgewbushlibrary.smu.edu/

Archives being moved in to George W. Bush library Houston Chronicle - 1 day ago

Peppard: George W. Bush signs on for Genesis Women's Shelter

doors to the public on May 1.

The George W. Bush Presidential Library and Museum in Dallas is set to open its

Here's your chance to support America's military families and take home a set of Presidential wheels. On Saturday night, President Bush's Ford F-150 King ...

President George W Bush and First Lady Laura Bush dance together during the ...

Home - The George W. Bush Presidential Library and Museum

### Google Knowledge Graph (started 2012)



Nicknames: 43. Bush 2.0. Bush II. Bush Jr., Dubva. Junior Shrub W

#### People also search for







Reagan

Feedback/More info

See results about



George H. W. Bush George Herbert Walker Bush is an

Romney

#### Schütze & Kessler: Semantic Search



- Entity-centric search works well for some searches.
- In other cases, we would like more expressive searches to take advantage of the semantics we have.
- We need to translate a user query to a formal representation (e.g., a SPARQL query).
- The user can help to disambiguate and correct if necessary.
- This has to be done in an interface that does not require the user to learn a formal language.

### Ontology-based Query Interpretation

- User enters query.
- The keywords in the query are mapped to ontology elements.
- Graph exploration is used to compute possible connections between the different ontology elements used in the query.
- The user can choose the interpretation closest to his intention.
- The interpretations are presented in a way that is intuitive for a user without knowledge about ontologies.

cepts Entities

### Example: Ask The Wiki (1)



**Step 1: Enter keywords**  $\longrightarrow$  **Step 2:** Choose interpretation  $\longrightarrow$  **Step 3:** View and refine results

- Enter keywords for your search
- For example you might enter mitarbeiter email, or projekt wissensmanagement kontaktperson.

	Semantic Wiki Search <sup>beta</sup>			
projekt wissensmanagement kontaktperson				
	Wiki Search!			
			J	

[http://www.aifb.kit.edu/web/Spezial:ATWSpecialSearch]

emantic Search

Keywords

ncepts Entities

Summar

### Example: Ask The Wiki (2)



Semantic Search

Keywords

Concepts

Entities

Summar

### Example: Ask The Wiki (3)



• Queries can be constructed out of a user query in natural language.

Entities

• Parsing the queries is a difficult problem – as you all know.

Example: Evi (1)

Entities

## time difference between Singapore and <u>Cebu</u>



- We can understand this question.
- We understood it as: What is the time difference between Singapore, officially the Republic of Singapore, the southeast Asian city-state off the southern tip of the Malay Peninsula, north of the equator and Cebu (Cebuano: Lalawigan sa Sugbo, ), the province in the Philippines, consisting of Cebu Island and 167 surrounding islands?
- We can answer this question.
- We have **0** user answers for this question, or its variants.
- We have **0** answers from other sites for this question, or its variants.

[http://www.evi.com]

Keywords

Concepts

Summary

Entities

### Example: Evi (2)

What is the time difference between Singa of Singapore, the southeast Asian city-stat Malay Peninsula, north of the equator and Sugbo, ), the province in the Philippines, co 167 surrounding Islands?	oore, offic e off the Cebu (Ce onsisting	cially the Rep southern tip buano: Lalaw of Cebu Islar	oublic of the vigan sa nd and
Analyse this question			
✓ Facts I used the following facts to provide this answered the following facts to provide the following facts to provide the factor of the facto	:	Show reason	ing 🔍
Singapore is having the relationship 'commonly	translate	s as' with "Sing disagree	japore" edit
Singapore is the time zone area for Singapore	agree	disagree	edit
Singapore Time has always been the time zone	for Singa agree	pore disagree	edit
+8 hours has been the time zone differential o August 20th 2010, 15:34:12	f Singapor	e Time since a	t least
"SGT" is the unique short form of Singapore Ti	ne	disagree	edit
Cebu is having the relationship 'commonly tran	islates as'	with "Cebu"	edit
Cebu is or has been in the Philippines		disagraa	odit

Keywords

Concepts

Summary

Entities

### Example: Evi (3)

What is the time difference between Singapore, officially the Republic of Singapore, the southeast Asian city-state off the southern tip of the Malay Peninsula, north of the equator and Cebu (Cebuano: Lalawigan sa Sugbo, ), the province in the Philippines, consisting of Cebu Island and 167 surrounding Islands?

#### Analyse this question

#### 🗸 Show facts

Reasoning 📿

I followed this chain of reasoning:

By calculation (generator: ["first\_parameter@trueknowledge.com"]) I know that: Singapore is the first parameter of Singapore and Cebu

By calculation (generator: ["second\_parameter@trueknowledge.com"]) I know that:

Cebu is the second parameter of Singapore and Cebu I know from locally stored knowledge that:

Singapore commonly translates as "Singapore" ([fact:

#### ["19590247@trueknowledge.com"]])

By calculation (generator: ["now@trueknowledge.com"]) I know that:

The current time is January 2nd 2013, 19:02:06

I know from locally stored knowledge that:

Singapore is the time zone area for Singapore ([fact:

#### ["98007601@trueknowledge.com"]])

Fact 1: Singapore Time is the time zone for Singapore ([fact:

#### ["20039199@trueknowledge.com"]])

Fact 1 is true for the earliest time when the fact can make sense onwards ([fact: ["20039200@trueknowledge.com"]]) Concepts Entities

### Example: Evi (4)



[http://www.evi.com]

#### Semantic Search

Keywords

Concepts Entities

### Querying as Problem Solving

- The core use case of the semantic web is describing a problem and searching for a solution by inferring one based on ontological knowledge.
- However, actual implementations of such are rare and they are usually quite simple, e.g.:
  - Wine Agent

(http://onto.stanford.edu:8080/wino/index.jsp): The user enters information on the flavors in a dish, and the system infers from the wine ontology a recommendation for a wine suitable to complement those flavors.

• Pizza Finder

(http://www.co-ode.org/downloads/pizzafinder/): The user specifies toppings he likes and dislikes on his pizza and the system infers from the pizza ontology a pizza fulfilling the restrictions.

### Example: Pizza Finder – Query

zza Finder		
The	e Manchester Piz	za Finder
	Included toppings:	
A SpicyTopping     MeatTopping		Add Rem
<ul> <li>DairyTopping</li> <li>CheeseTopping</li> <li>EggTopping</li> <li>VegetableTopping</li> <li>SauceTopping</li> </ul>	DairyTopping	
<ul> <li>► A Full opping</li> <li>► A NutTopping</li> <li>► A HerbSpiceTopping</li> <li>► FishTopping</li> </ul>	CExcluded toppings:	Add Rem

[http://www.co-ode.org/downloads/pizzafinder/]

Entities

### Example: Pizza Finder – Result



[http://www.co-ode.org/downloads/pizzafinder/]

The Semantic Web	Keywords	Concepts	Summary	
Outline				

- 1 The Semantic Web
- 2 Semantic Search
- Inhancing Keyword Search
- 4 Searching for Concepts and Relations
- 5 Beyond Document Retrieval



cepts Entities

### Query Construction

### Task Definition [HHLV09]

Task: "Assisting the user in the specification of the information need."

Result: "A representation of the information need in terms of primitives of a language supported by the system."

- Augmenting traditional keyword search. Examples: CleverSearch, Ask The Wiki
- Select a class and restrict property values. Examples: Yummly
- Natural language queries. Examples: Evi
- Querying as problem solving.
   Examples: Wine Agent, Pizza Finder

Keywords

oncepts Entities

### Query Processing

### Task Definition [HHLV09]

Task: "Matching the user information need as specified in the user query against the system resource." Result: "A set of documents [or triples] containing the information

that satisfy the user need."

 Query processing on a keyword index with ontology-based query expansion.

Examples: CleverSearch

- Query processing on several keyword indices based on fields defined by the ontology.
   Examples: Yummly?
- Query Processing on a Semantic Repository. Examples: Ask The Wiki, Yummly?, Wine Agent, Pizza Finder, Evi

Entities

### **Result Presentation**

### Task Definition

Task: "Present the search results to the user." Result: "A user-friendly, easy to navigate presentation of information."

- Present a ranked list of documents. Examples: CleverSearch, Yummly
- Present a ranked list of entities and for each entity all known information about it.
   Examples: Wine Agent, Pizza Finder, SWSE
- Present a structured table of possible elements fulfilling the query restrictions (tuple set).
   Examples: Ask The Wiki

• . . .

- Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph and York Sure. Semantic Web. Grundlagen. Springer textbook, 2008. (General introduction to semantic web)
- Pascal Hitzler, Markus Krötzsch and Sebastian Rudolph. Foundations of Semantic Web Technologies. Chapman & Hall/CRC, 2009. (General introduction to semantic web)
- Sudhir Agarwal, Stephan Grimm, Peter Haase, Pascal Hitzler, and Denny Vrandecic. Semantic Search and Information Integration. Slides for Course Semantic Web Technologies II, Universität Karlsruhe, SS 2009.

### Specific Literature (1)

- Peter Haase, Daniel M. Herzig, Mark Musen and Duc Thanh Tran. Technical Report: Semantic Wiki Search. Technical Report, Universität Karlsruhe, 2008. http://www.aifb.kit.edu/web/Spezial:ATWSpecialSearch
- Peter M. Kruse, André Naujoks, Dietmar Rösner and Manuela Kunze. Clever search: A wordnet based wrapper for internet search engines. In: Proceedings of the 2nd GermaNet Workshop, 2005. http://wdok.cs.uni-magdeburg.de/clever-search/
- http://www.co-ode.org/downloads/pizzafinder/



- Aidan Hogan, Andreas Harth, Jürgen Umbrich, Sheila Kinsella, Axel Polleres and Stefan Decker. Searching and Browsing Linked Data with SWSE: the Semantic Web Search Engine. In: Journal of Web Semantics, Volume 9, Issue 4, December 2011. http://swse.deri.org/
- Eric Hsu and Deborah McGuinness. Wine Agent: Semantic Web Testbed Application. Workshop on Description Logics, 2003. http://onto.stanford.edu:8080/wino/index.jsp

http://www.yummly.com/blog/about/ (Yummly blog)
http://www.yummly.com/recipes