Semantic Technologies for Advanced Medical Image and Information Search

THESEUS-MEDICO

Pinar Wennerberg
pinar.wennerberg.ext@siemens.com

Vienna, 26.09.2008

Agenda

- Introduction
  - THESEUS-MEDICO
- Knowledge Engineering in the Medical Domain
  - Domain Specific Knowledge Engineering Challenges
  - Requirements Analysis
  - The KEMM Methodology
  - Query Pattern Derivation
  - Ontology Identification
  - Medical Ontology Alignment
- Final Remarks

Introduction
THESEUS-MEDICO

Problem:

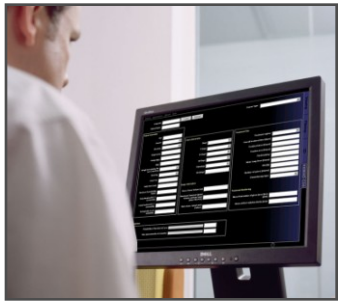
Many individual medical applications, but no common semantics

BUT:

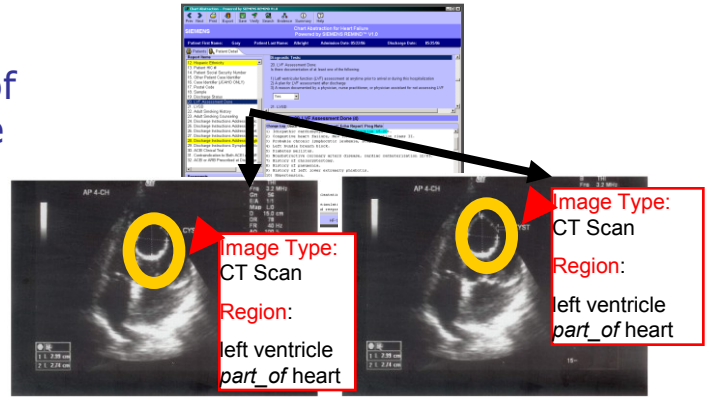
Queries are not arbitrary, instead based on anatomy, physiology, pathology…

- e.g. only heart has left ventricle
- e.g. certain spatial relations between the organ regions/segments

Proposed solution:



"Show me the CT scans & records of patients with an enlargement in the dimension of the lymph node in the neck"



Domain Specific Knowledge Engineering Challenges

THESEUS MEDICO

- Challenges of knowledge engineering in the medical domain

  - Communication based
    - Knowledge acquisition bottleneck or knowledge elicitation

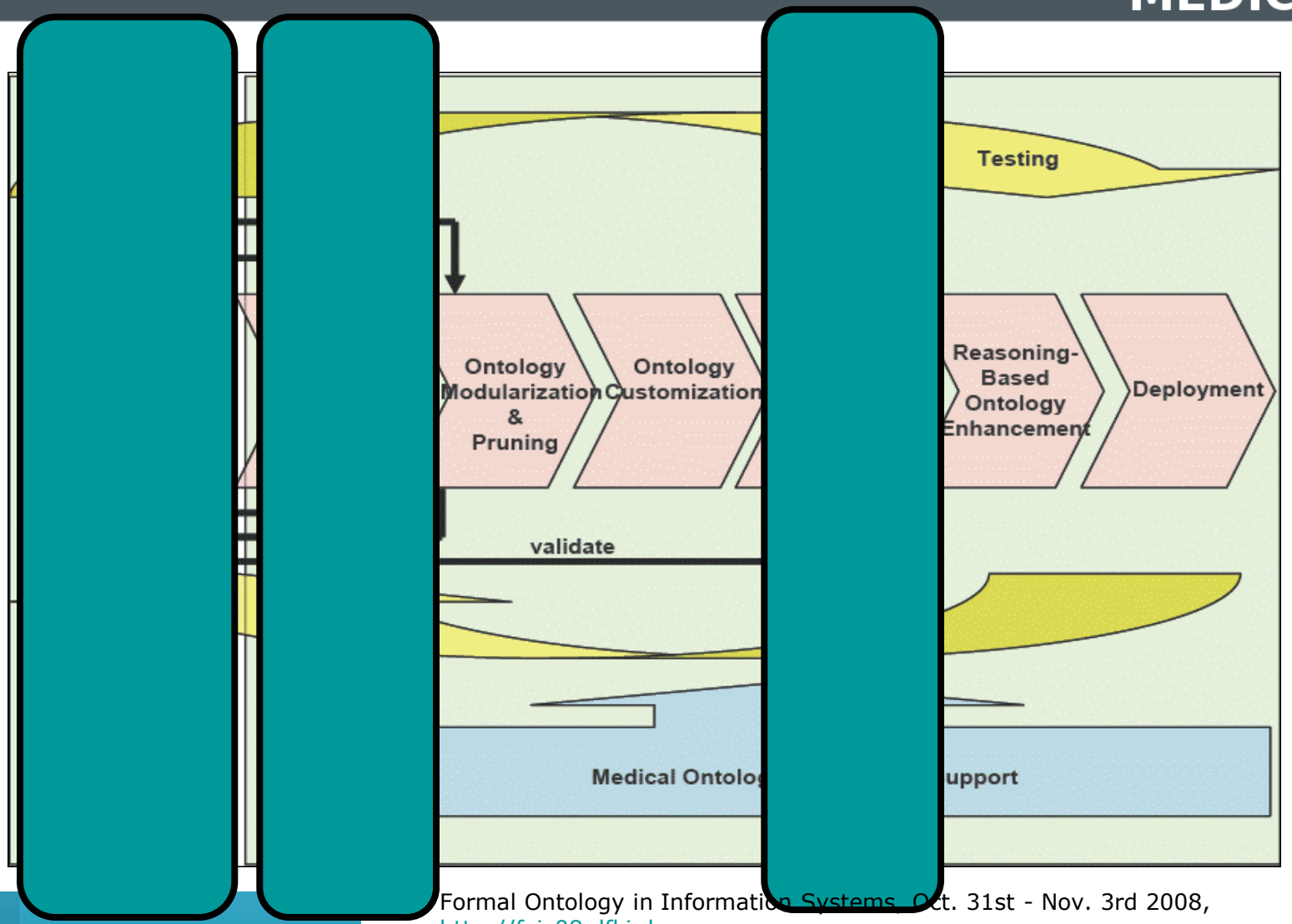
  - Medical ontology engineering based
    - Large and comprehensive medical ontologies
    - Knowledge is opaque to the engineer due to special vocabulary
    - Complex – several hierarchies, many relationships
    - Sensitive domain, knowledge has to be accurate
    - Modelling challenge due to lack of domain expertise

Requirements Analysis

To achieve the objectives set for MEDICO in a systematic way we identified the following requirements:

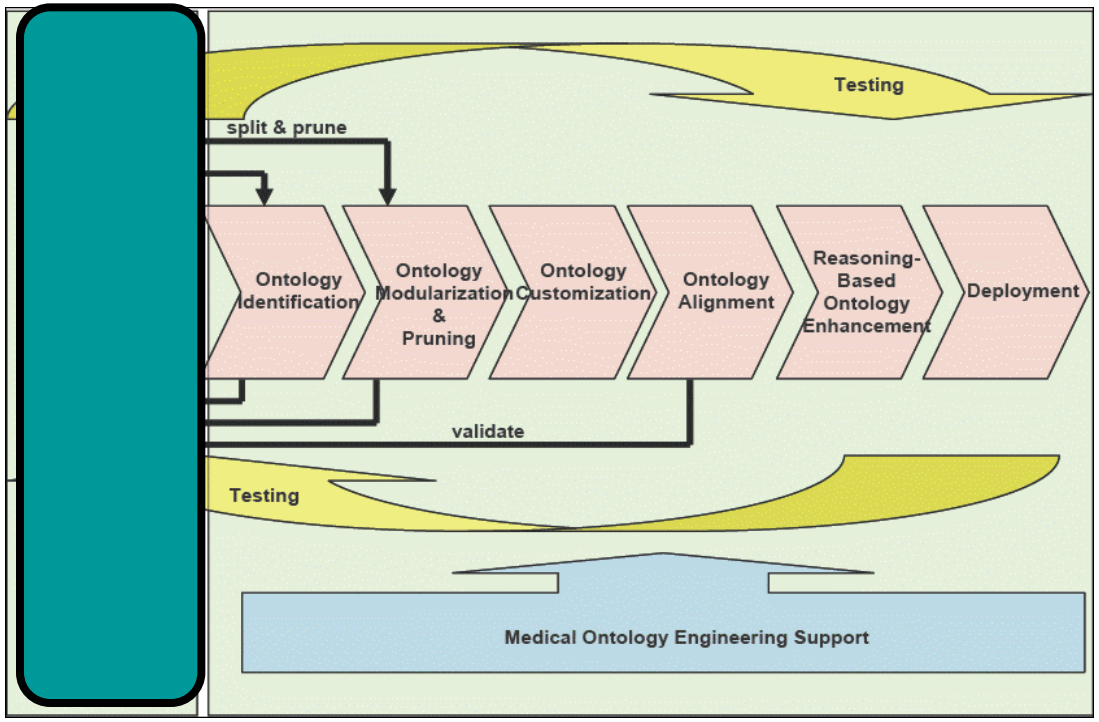
- Query Pattern Derivation
- Ontology Identification
- Ontology Modularization and Pruning
- Ontology Customization
- Ontology Alignment
- Reasoning-Based Ontology Enhancement
- Testing and Deployment

KEMM: Knowledge Engineering Methodology in the Medical Domain

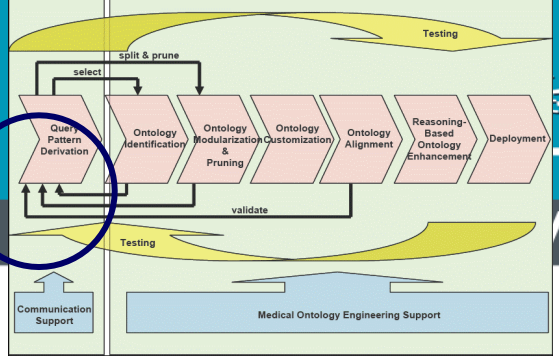


Formal Ontology in Information Systems, Oct. 31st - Nov. 3rd 2008, http://fois08.dfki.de

Query Pattern Derivation



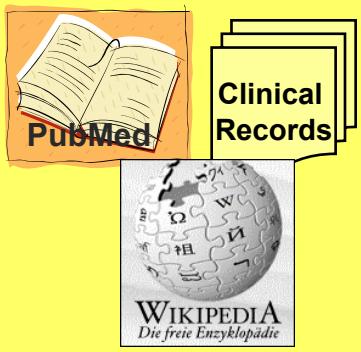
Query Pattern Derivation



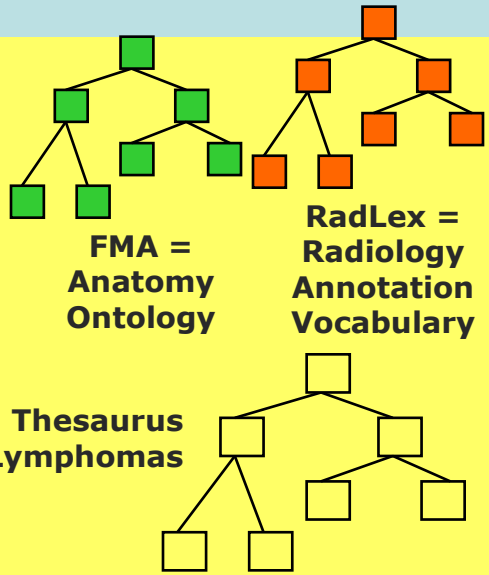
THESEUS MEDICO

Challenge:
„What are typical queries clinicians or radiologists are interested in?"

we use….



&



…to achieve:

„Enlargement in the dimension of the lymph node in the neck?"

<Anatomical Structure>		<Anatomical Structure>
Lymph-Node	located-in	Neck
Neck	has-dimension	XYZ

joint corpora of anatomy, disease, radiology

…. for identifying the most relevant concepts & relationships

…. derive query patterns

Clinical Evaluation

Data Processing: RadLex & FMA

THESEUS MEDICO

Anatomy Corpus

	Term	FMA	RadLex	ICD	Freq.	Score	PartOfSpeech
2	lateral	y	y	n	464	338724,00	JJ
3	anterior	y	y	n	452	314721,00	JJ
4	artery	y	y	n	237	281961,00	NN
5	anterior spinal artery	y	y	n	2	219894,33	JJ JJ NN
6	lateral thoracic artery	y	y	n	2	217815,33	JJ JJ NN
7	cortex	y	y	n	359	215296,00	NN
8	anterior nucleus	n	y	n	6	211375,09	JJ NN
9	anterior cerebral artery	y	y	n	2	208527,33	JJ JJ NN
10	lateral plantar artery	y	y	n	1	207540,33	JJ JJ NN
11	lateral sacral artery	n	y	n	5	207071,33	JJ JJ NN
12	lateral posterior nucleus	n	y	n	3	206880,73	JJ JJ NN
13	anterior tibial artery	n	y	n	8	199866,00	JJ JJ NN
14	anterior cecal artery	y	n	n	1	198894,33	JJ JJ NN
15	anterior choroidal artery	n	y	n	1	198894,33	JJ JJ NN
16	anterior communicating artery	n	y	n	2	198894,33	JJ VBG NN
17	lateral dorsal nucleus	n	y	n	3	188592,73	JJ JJ NN
18	first	n	y	n	388	176400,00	JJ
19	anterior superior alveolar artery	n	y	n	1	175190,75	JJ JJ JJ NN
20	lateral ventricle	n	y	n	7	174564,00	JJ NN

Radiology Corpus

	Term	FMA	RadLex	ICD	Freq.	Score	PartOfSpeech
2	x-ray	n	y	n	253	81901,64	NN
3	imaging modality	n	y	n	6	58682,00	NN NN
4	volume imaging	n	y	n	1	57855,09	NN NN
5	molecular imaging	n	y	n	4	57850,00	JJ NN
6	mr imaging	n	y	n	9	57850,00	JJ NN
7	magnetic resonance imaging	n	y	n	44	48072,67	JJ NN NN
8	nuclear medicine imaging	n	y	n	6	43438,97	JJ NN NN
9	functional magnetic resonance imaging	n	y	n	5	36279,50	JJ JJ NN NN
10	dual energy x-ray absorptiometry	n	y	n	2	20562,47	JJ NN NN NN
11	ultrasound	n	y	n	114	15376,00	NN
12	radiation dose	n	y	n	14	12168,97	NN NN
13	magnetic resonance angiography	n	y	n	3	10046,33	JJ NN NN
14	magnetic resonance spectroscopy	n	y	n	10	9659,67	JJ NN NN
15	small	n	y	n	90	8649,00	JJ
16	nuclear	n	y	n	80	7921,00	JJ
17	3d ultrasound	n	y	n	1	7688,00	CD NN
18	first	n	y	n	83	7056,00	JJ
19	artery	y	y	n	65	6724,00	NN
20	computed tomography	n	y	n	42	5860,00	JJ NN

Steps:

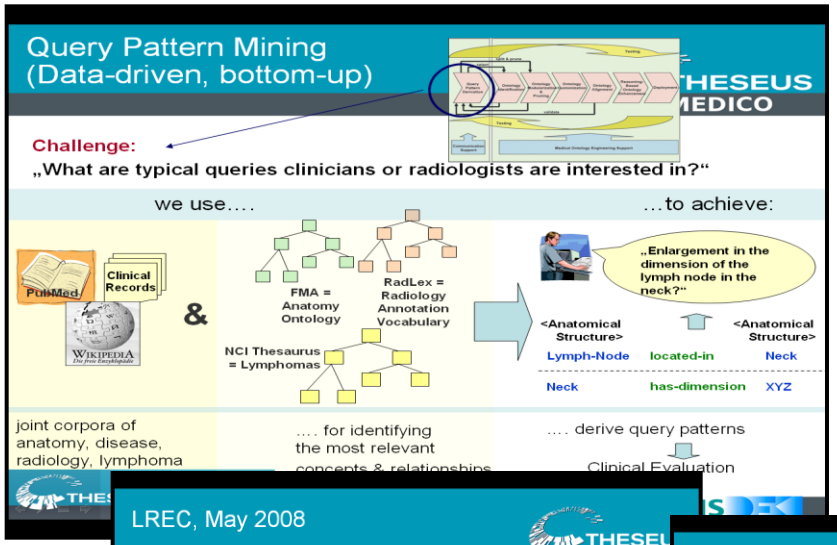
- all text sections of each corpus through the TnT part-of-speech parser (Brants, 2000)
- extract all nouns in the corpus
  - compute a relevance score (chi-square) for each
  - by comparing anatomy & radiology frequencies respectively with those in the British National Corpus

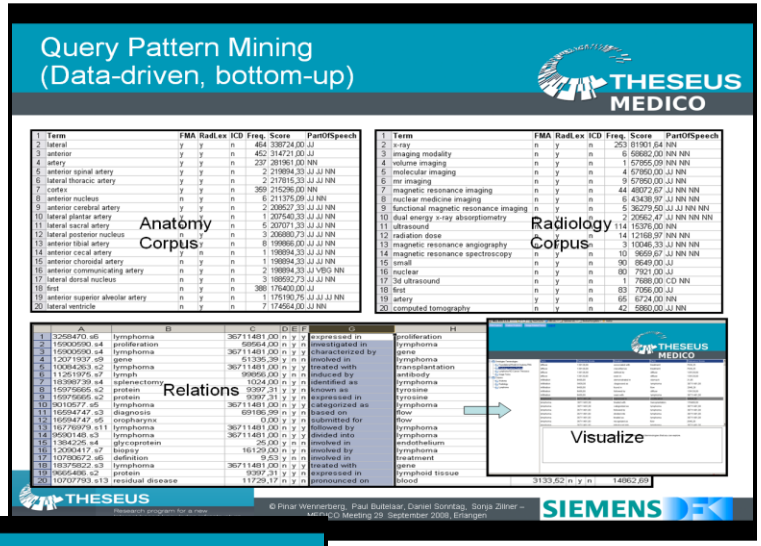
Next:

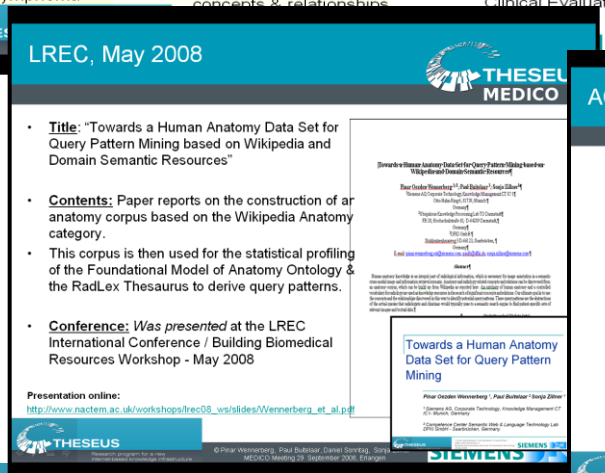
- parse all sentences in all corpora and annotate them with predicate-structure information

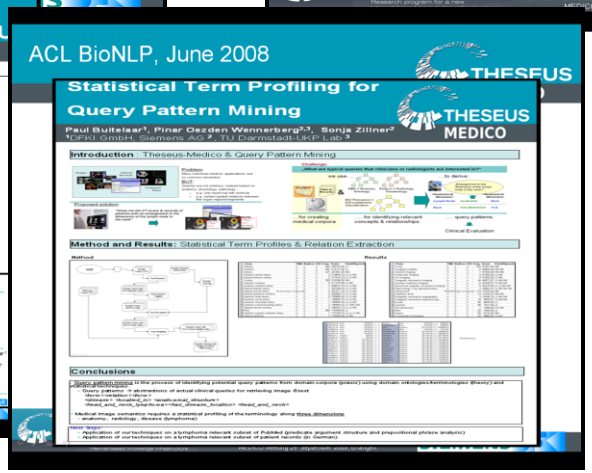
Query Pattern Mining & Dissemination

- Identified domain ontologies and set up corpora
- Extracted domain terms and relations





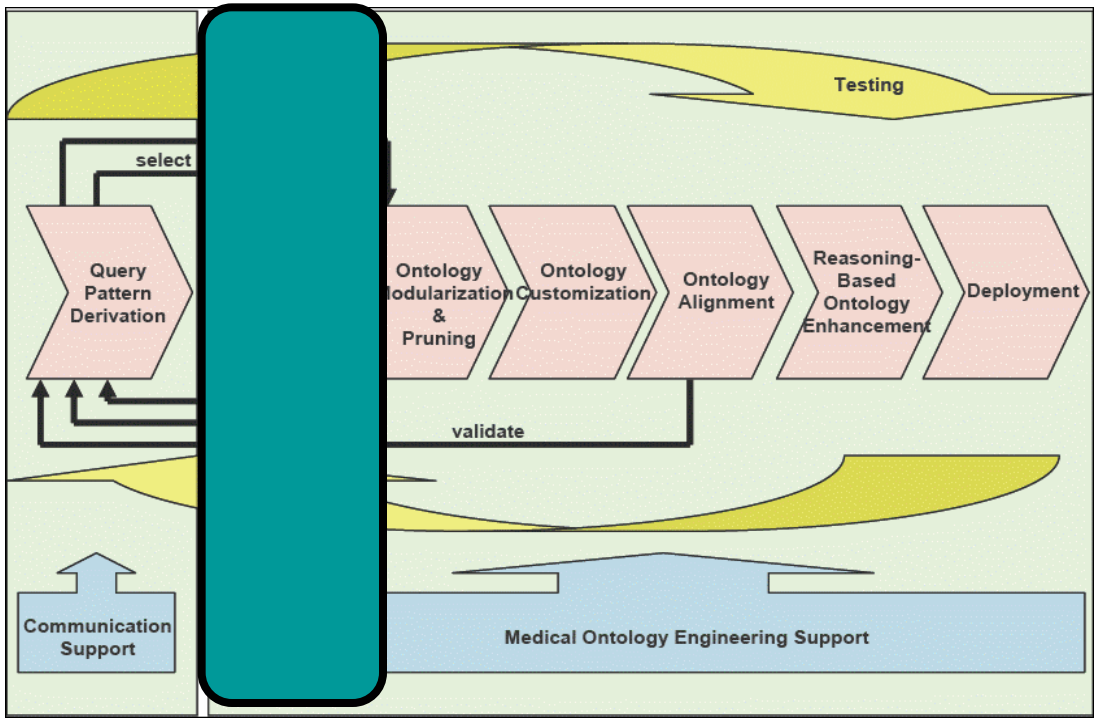




Dissemination:

- Published and presented in international conferences
- Patent pending

Ontology Identification



Ontology Identification

Criteria for selecting MEDICO relevant ontologies and thesauri

- Three perspectives
  - the anatomical spatial: body parts and their locations,
  - the radiology: relationships between various image modalities and anatomical regions as shown on medical images,
  - the disease-pathology: distinction between normal and abnormal imaging features
- Representation Language
  - as predefined by MEDICO either in OWL or easily transferable
- Comprehensiveness
  - most comprehensive (e. g., FMA) to avoid missing relevant important information
- Popularity: being non-experts, popularity is our guide…
  - how well known
  - how well documented
  - how many projects using the ontology
- Semantic Formalism
  - Description Logics as predefined by THESEUS-MEDICO

Ontology Identification

THESEUS MEDICO

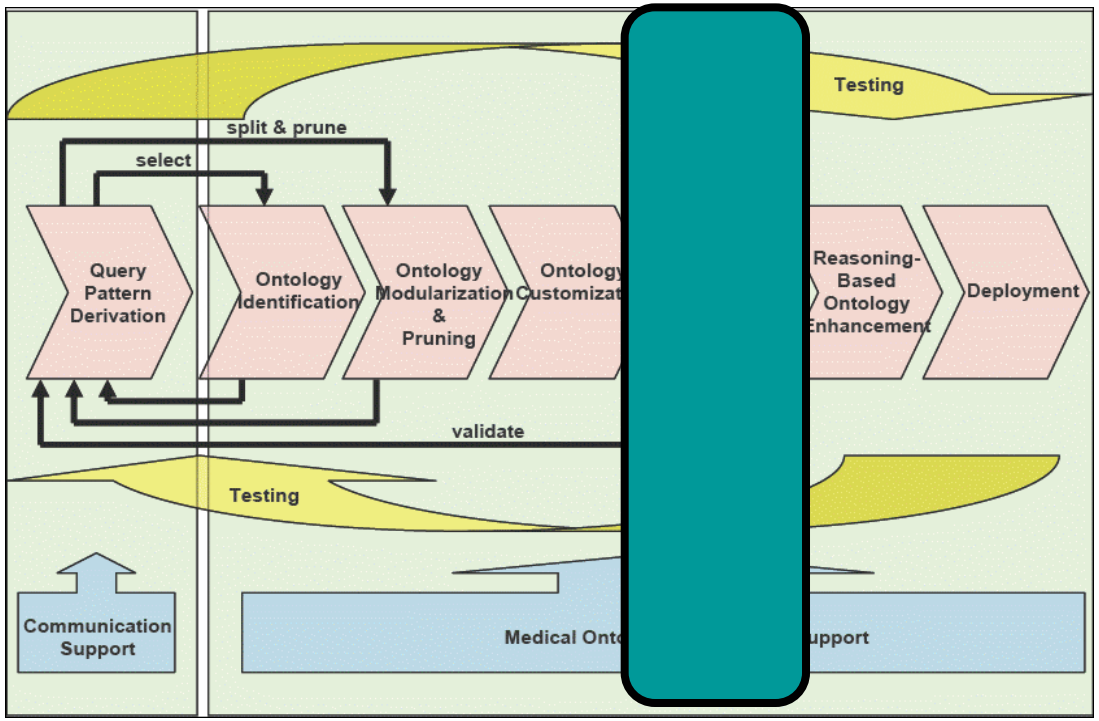
Relevant domain ontologies for MEDICO

Ontology	Perspective	# of Concepts	# of Rel. Types	# of Syns.	Format	Data File
FMA	anatomy	71.202	170	28.488 (up to 6 per concept)	OWL, Frames, OBO*	http://depts.washington.edu/ventures/UW_Technology/Express_Licenses/FMA.php
RadLex	radiology	11.962	14		XML, Frames	ftp://ftp.ihe.net/RadLex
NCI Cancer Thesaurus	disease	34.000	182		OWL, Text, OBO	ftp://ftp1.nci.nih.gov/pub/cacore/EVS/NCI_Thesaurus/

* OBO Foundry

Ontology Alignment –

Current Focus &
Work in Progress



Medical Ontology Alignment

- Often, there is a need to use the knowledge from multiple ontologies

→ Thus we need to integrate knowledge from disparate domain ontologies about anatomy, radiology and disease/pathology (i.e. the three perspectives)

- Semantic Integration → Ontology Alignment → Identify equivalent concepts (and relations) from disparate domain ontologies

MEDICO-KEMM Approach to Ontology Alignment

- Leveraging on the current technologies
  - Hybrid (Schema based and Instance based)
- Combinatory
  - Linguistic-based (transformation rules and formal grammers)
  - Corpus-based (observation of co-occurrences)
  - Dialogue-based (interaction with user, relevance feedback)

Final Remarks

Final Remarks

- Ontology engineering in the medical domain is challenging not only because of technical issues but also because of communication mismatches.
- Medical ontologies have specific characteristics (large, complex, opaque semantics…), which requires special treatment.
- Expert knowledge is not always available in the medical domain because the doctors are almost always in lack of time. This requires finding alternative solutions to acquire expert knowledge.
- Medical image semantics requires integration of information across three dimensions; anatomy, radiology and diseases, pathology.
- Clinicians/radiologists are not aware of the technologies available to them. One important step is to inform them about all what they could demand.

Final Remarks

THESEUS MEDICO

fois2008

5th international conference on formal ontology in information systems

HOME | ORGANIZERS | DATES | SUBMISSIONS | PAPERS | REGISTRATION | VEN

Welcome to FOIS 2008, the 5th International Conference on Formal Ontology in Information Systems

Saarbrücken, Germany
Oct 31st - Nov 3rd 2008
co-located with ISWC 2008, Karlsruhe, Germany *(Oct 26 - 30)*

Registration is now open!

Since its inception ten years ago, the International Conference on Formal Ontology in Information Systems has explored the multiple perspectives on the notion of ontology that have arisen from such diverse research communities as philosophy, logic, computer science, cognitive science, linguistics, and various scientific domains.

As ontologies have been applied in new and exciting domains such as the World Wide Web, bioinformatics, and geographical information systems, it has become evident that there is a need for ontologies that have been developed with solid theoretical foundations based on philosophical, linguistic, and logical analysis. Similarly, there is also a need for theoretical research that is driven by the issues that have been raised by recent work in the more applied domains.

FOIS is intended to be a forum in which to explore this interplay between the theoretical insights of formal ontology and their application to information systems and emerging semantic technologies.

- Submissions
- Accepted Papers
- Venue
- Registration
- Program
- News
- Printable
  - Flyer
  - Conference Booklet

Cases on Semantic Interoperability for Information Systems Integration: Practices and Applications
A book edited by Dr. Yannis Kalfoglou, Ricoh Europe Plc, UK

Introduction

Semantic interoperability and integration is concerned with the use of explicit semantic descriptions to facilitate information and systems integration. Semantic technologies provide the means to attach meaning to conventional concepts which makes it possible to automatically process and integrate large amounts of information without human intervention.

The Overall Objective of the Book

This case book will be a concise and elaborate analysis of the state-of-the-art in the emergent field of semantic interoperability and integration. It will include a number of cases covering the need for the use of semantics to enable intelligent information integration in organizational and personal computing settings. The case book will explain in depth, the issues involved with integrating large amounts of heterogeneous information, and point to the deficiencies of current systems. It will complement existing publications in this field and provide the reader with a clear understanding of why semantics are needed and how to make use of them in information integration tasks.

The Target Audience

Research practitioners in academia and industry, research students, information and business analysts, information management professionals, engineers, and software developers. The book will assist in practitioners' quest for knowledge on applying semantic technologies to their information management tasks; research students and associates will find this resource as a one stop reference resource for the semantic integration and interoperability field; information management practitioners will find this resource a valuable aid in their task on improving everyday practice; engineers and software developers will find this resource a repository of cases where detailed engineering procedures and methods are exposed for the benefit of replicating the solution to similar settings.

Recommended topics include, but are not limited to, the following:

- Use of ontology mapping technology to semantic interoperability
- Use of semantic technology in Enterprise Application Integration (EAI) settings
- Use of semantic technology for Intelligent Information Integration (III) applications
- Cost of integration using semantic technologies lessons learnt from applying and/or deploying semantic technologies for integration
- Scope and benefits of semantic technologies for integration
- Theoretical underpinnings of semantic integration
- Impact of Semantic Web and Web 2.0 to integration solutions

Book Chapter:
"Cases on Semantic Interoperability"

Disseminating MEDICO

Final Remarks

- We would like to thank our project partners:
  - DFKI Saarbrücken Language Technology Lab
  - DFKI Kaiserslautern Knowledge Management
  - University of Erlangen

  for their valuable contributions…

Contact

Pinar Wennerberg
pinar.wennerberg.ext@siemens.com

Dr. Sonja Zillner
sonja.zillner@siemens.com

Siemens AG
Corporate Technology
Knowledge Management, CT IC 1
Otto-Hahn-Ring 6, 81739
München

Thank you!