COMP718: Ontologies and Knowledge Bases

Lecture 1: Introduction to Knowledge bases, ontologies, and the Semantic Web

Maria Keet
email: keet@ukzn.ac.za
home: http://www.meteck.org

School of Mathematics, Statistics, and Computer Science
University of KwaZulu-Natal, South Africa

07 February 2012

Outline

1. Administrivia
2. Introduction
3. Use case: the Semantic Web
4. What is an Ontology?

Outline

1. Administrivia
2. Introduction
3. Use case: the Semantic Web
4. What is an Ontology?

Administrivia (1/2)

- This course consists of 13 lectures, exercises, a written assignment, a practical assignment, and a mini-project in small groups (see course outline and LN for details)
- Continuous assessment: practical assignment [20%], mini-project [30%], written test [50%]. You have to submit something for each one to have a chance to pass the course.
- Each lecture takes about 2-2.5 hours, labs&self study 6-8 hours/week
- Following the lectures will be easier when you have read the recommended and required reading beforehand

Administrivia (1/2)

- This course consists of 13 lectures, exercises, a written assignment, a practical assignment, and a mini-project in small groups (see course outline and LN for details)
- Continuous assessment: practical assignment [20%], mini-project [30%], written test [50%]. You have to submit something for each one to have a chance to pass the course.
- Each lecture takes about 2-2.5 hours, labs&self study 6-8 hours/week
- Following the lectures will be easier when you have read the recommended and required reading beforehand

Administrivia Introduction Use case: the Semantic Web What is an Ontology? Summary

Administrivia (1/2)

- This course consists of 13 lectures, exercises, a written assignment, a practical assignment, and a mini-project in small groups (see course outline and LN for details)
- Continuous assessment: practical assignment [20%], mini-project [30%], written test [50%]. You have to submit something for each one to have a chance to pass the course.
- Each lecture takes about 2-2.5 hours, labs&self study 6-8 hours/week
- Following the lectures will be easier when you have read the recommended and required reading beforehand

4/26

Administrivia (2/2)

- Slides, additional reading material, and answers to the exercises will be made available through the module's Moodle page
  - **The slides serve as a teaching aid, not as a neat summary**
- The topics covered in this course are of an introductory nature and only a selection of core and elective topics will be addressed; *this is an active research field...*.
- ... so there is no single textbook (yet) that covers all topics for the novice ontologist, has exercises with given, clear answers etc, but...
- ... there are new lecture notes (though you still have to learn to read some scientific literature)

Administrivia Introduction Use case: the Semantic Web What is an Ontology? Summary

Administrivia (2/2)

- Slides, additional reading material, and answers to the exercises will be made available through the module's Moodle page
  - The slides serve as a teaching aid, not as a neat summary
- The topics covered in this course are of an introductory nature and only a selection of core and elective topics will be addressed; this is an active research field....
- ... so there is no single textbook (yet) that covers all topics for the novice ontologist, has exercises with given, clear answers etc, but...
- ... there are new lecture notes (though you still have to learn to read some scientific literature)

5/26

Administrivia (2/2)

- Slides, additional reading material, and answers to the exercises will be made available through the module's Moodle page
  - **The slides serve as a teaching aid, not as a neat summary**
- The topics covered in this course are of an introductory nature and only a selection of core and elective topics will be addressed; *this is an active research field....*
- ... so there is no single textbook (yet) that covers all topics for the novice ontologist, has exercises with given, clear answers etc, but...
- ... there are new lecture notes (though you still have to learn to read some scientific literature)

Outline

1 Administrivia
2 Introduction
3 Use case: the Semantic Web
4 What is an Ontology?

Databases vs. Knowledge bases

- Main differences:
  - Representation of the knowledge
  - Rules
  - Reasoning to infer new or implicit knowledge, detect inconsistencies of the knowledge base
  - Open World Assumption (vs. Closed World Assumption)
- Some references:
  - Hopgood, A.A. *Intelligent systems for engineers and scientists*. CRC Press: Boca Raton, Florida, USA, 2nd ed. 2000.
  - Any PROLOG book, and, e.g., LPA's software *flex*
  - FLORA (based on F-logic)
  - The Description Logic Handbook

Databases vs. Knowledge bases

- Main differences:
  - Representation of the knowledge
  - Rules
  - Reasoning to infer new or implicit knowledge, detect inconsistencies of the knowledge base
  - Open World Assumption (vs. Closed World Assumption)
- Some references:
  - Hopgood, A.A. *Intelligent systems for engineers and scientists*. CRC Press: Boca Raton, Florida, USA, 2nd ed. 2000.
  - Any PROLOG book, and, e.g., LPA's software *flex*
  - FLORA (based on F-logic)
  - The Description Logic Handbook

Conceptual data models vs ontologies

- Main differences:
  - Information needs for one application vs. representing the knowledge of a subject domain (regardless the particular application)
  - Formalization in a logic language (though one could do that for conceptual models as well)
- An ontology as a layer on top of conceptual data models
  - To improve the quality of a conceptual data model (hence, the software)
  - To facilitate database integration, or prevent the usual data integration problems

Administrivia | Introduction | Use case: the Semantic Web | What is an Ontology? | Summary

Conceptual data models vs ontologies

- Main differences:
  - Information needs for one application vs. representing the knowledge of a subject domain (regardless the particular application)
  - Formalization in a logic language (though one could do that for conceptual models as well)
- An ontology as a layer on top of conceptual data models
  - To improve the quality of a conceptual data model (hence, the software)
  - To facilitate database integration, or prevent the usual data integration problems

8/26

What is the usefulness of an ontology?

- Making, more or less precisely, the (dis-)agreement among people explicit
- Enrich software applications with the additional semantics
- Thus, practically, improving computer-computer, computer-human, and human-human communication

Examples using different features

- Data(base) integration
- Instance classification
- Matchmaking and services
- Querying, information retrieval
  - Ontology-Based Data Access
  - Ontologies to improve NLP
- Bringing more quality criteria into conceptual data modelling to develop a better model (hence, a better quality software system)
- Orchestrating the components in semantic scientific workflows, e-learning, etc.

Some diverse applications

- Deep QA with Watson; uses over 100 techniques, including ontologies for integration[1]
- Data mining management; see e.g., the e-LICO project's Data Mining and OPtimization ontology and infrastructure for classification, http://www.e-lico.eu and http://www.dmo-foundry.org
- Taverna for semantic scientific workflows, integration (domain and workflow ontologies), consistency checking, taxonomic classification, http://taverna.sourceforge.net/

[1] An entertaining presentation was given by one of the algorithm developers, Chris Welty, at ESWC'11, providing funny examples when this went wrong, why, an what they learned from it to improve on the algorithms. It is available as video lecture at http://videolectures.net/eswc2011_welty_watson/

Outline

1. Administrivia
2. Introduction
3. Use case: the Semantic Web
4. What is an Ontology?

Introduction (some motivations for ontologies and knowledge bases)

- AI put to the test in the (uncontrollable?) very large field
- Adding meaning to plain HTML pages and Web 2.0 by using theory and technologies of KBs and ontologies
  - *But there is more to ontologies and knowledge bases than their application in the Semantic Web!*
- See slides semweb-intro.pdf

Outline

1. Administrivia
2. Introduction
3. Use case: the Semantic Web
4. What is an Ontology?

Background

– Aristotle and colleagues: **O**ntology
– Engineering: ontolog**ies** (count noun)

– Investigating reality, representing it
– Putting an engineering artifact to use

What then, is this engineering artifact?

(Guarino, 2002)

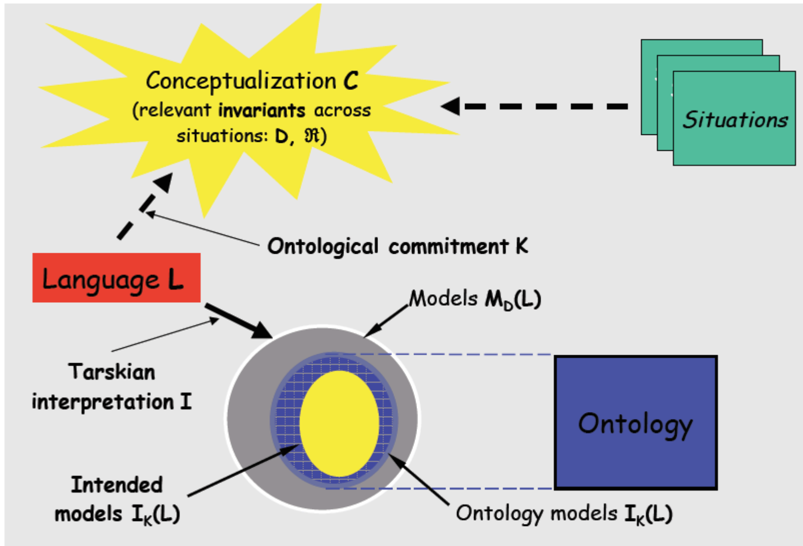
A few definitions

- Most quoted: "An ontology is a specification of a conceptualization" (by Tom Gruber, 1993)
- "a formal specification of a shared conceptualization" (by Borst, 1997)
- "An ontology is a formal, explicit specification of a shared conceptualization" (Studer et al., 1998)
- What is a *conceptualization*, and a *formal, explicit specification*? Why *shared*?

A few definitions

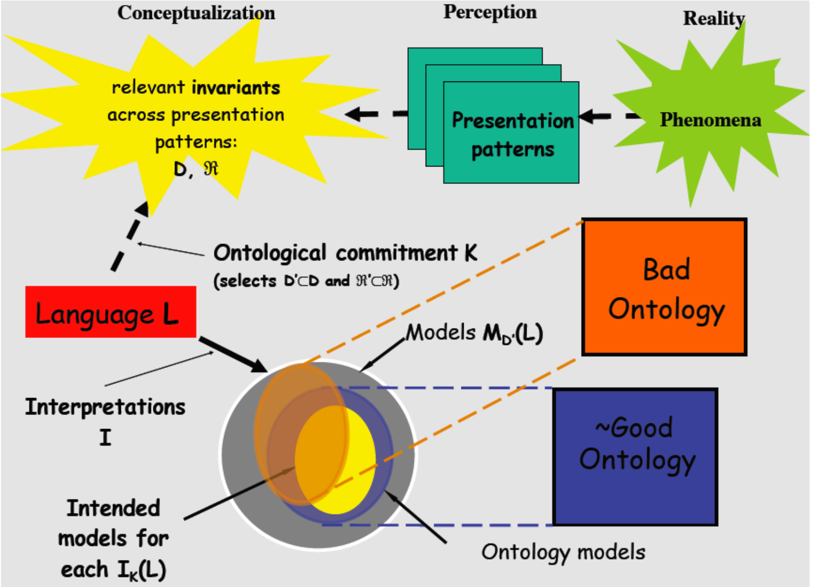
- Most quoted: "An ontology is a specification of a conceptualization" (by Tom Gruber, 1993)
- "a formal specification of a shared conceptualization" (by Borst, 1997)
- "An ontology is a formal, explicit specification of a shared conceptualization" (Studer et al., 1998)
- What is a *conceptualization*, and a *formal, explicit specification*? Why *shared*?

Ontologies and meaning

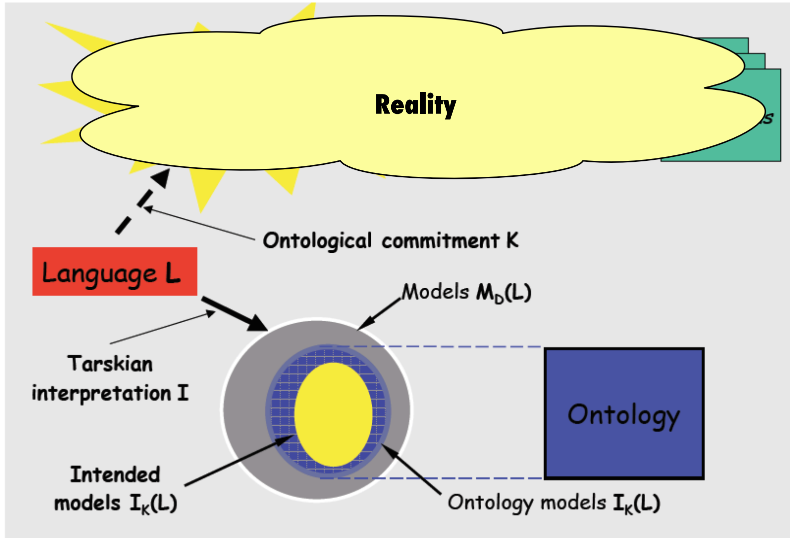


(Guarino, 2002)

Ontologies and meaning



Ontologies and reality



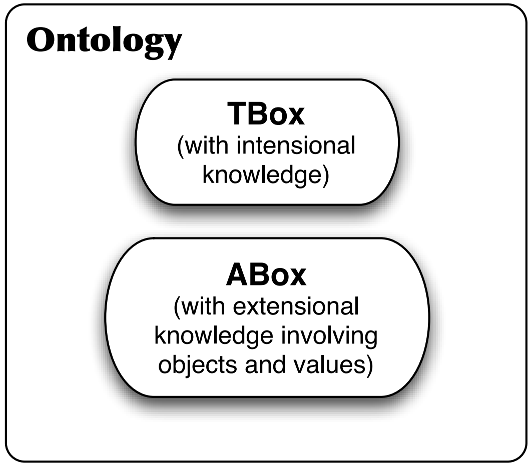
More definitions

- More detailed: "An ontology is a logical theory accounting for the *intended meaning* of a formal vocabulary, i.e. its *ontological commitment* to a particular *conceptualization* of the world. The intended models of a logical language using such a vocabulary are constrained by its ontological commitment. An ontology indirectly reflects this commitment (and the underlying conceptualization) by approximating these intended models." (Guarino, 1998)
- And back to a simpler definition: "with an ontology being equivalent to a Description Logic knowledge base" (Horrocks et al, 2003)

More definitions

- More detailed: "An ontology is a logical theory accounting for the *intended meaning* of a formal vocabulary, i.e. its *ontological commitment* to a particular *conceptualization* of the world. The intended models of a logical language using such a vocabulary are constrained by its ontological commitment. An ontology indirectly reflects this commitment (and the underlying conceptualization) by approximating these intended models." (Guarino, 1998)
- And back to a simpler definition: "with an ontology being equivalent to a Description Logic knowledge base" (Horrocks et al, 2003)

Description Logic knowledge base



From logical to ontological level

- Logical level (no structure, no constrained meaning[2]):
  - $\exists x(Apple(x) \land Green(x))$

adapted from (Guarino, 2008)

[2] meaning in the sense of subject domain semantics, not formal semantics

From logical to ontological level

- Logical level (no structure, no constrained meaning[2]):
  - $\exists x(Apple(x) \land Green(x))$
- Epistemological level (structure, no constrained meaning):
  - $\exists x : apple\ Green(x)$ (many-sorted logics)
  - ~~$\exists x : green\ Apple(x)$~~
  - $Apple(a)$ and $hasColor(a, green)$ (description logics[3])
  - ~~$Green(a)$ and $hasShape(a, apple)$~~

adapted from (Guarino, 2008)

[2] meaning in the sense of subject domain semantics, not formal semantics

[3] DL has a model-theoretic semantics, so the axioms have a meaning in that sense of 'meaning/semantics'

From logical to ontological level

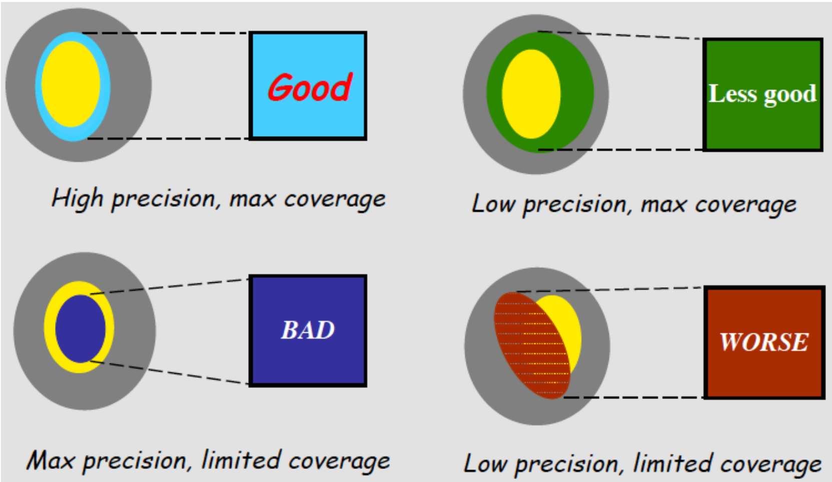
- Logical level (no structure, no constrained meaning[2]):
  - $\exists x(Apple(x) \land Green(x))$
- Epistemological level (structure, no constrained meaning):
  - $\exists x : apple\ Green(x)$ (many-sorted logics)
  - ~~$\exists x : green\ Apple(x)$~~
  - $Apple(a)$ and $hasColor(a, green)$ (description logics[3])
  - ~~$Green(a)$ and $hasShape(a, apple)$~~
- Ontological level (structure, constrained meaning):
  - Some structuring choices are excluded because of ontological constraints
  - *Apple* carries an identity condition (and is a sortal), *Green* does not (is a qualia ['value'] of the quality ['attribute'] *hasColor* that a thing has)

adapted from (Guarino, 2008)

[2] meaning in the sense of subject domain semantics, not formal semantics

[3] DL has a model-theoretic semantics, so the axioms have a meaning in that sense of 'meaning/semantics'

Quality of the ontology



(Guarino, 2002)

Initial Ontology Dimensions that have Evolved

- Semantic
  - Degree of Formality and Structure
  - Expressiveness of the Knowledge Representation Language
  - Representational Granularity
- Pragmatic
  - Intended Use
  - Role of Automated Reasoning
  - Descriptive vs. Prescriptive
  - Design Methodology
  - Governance

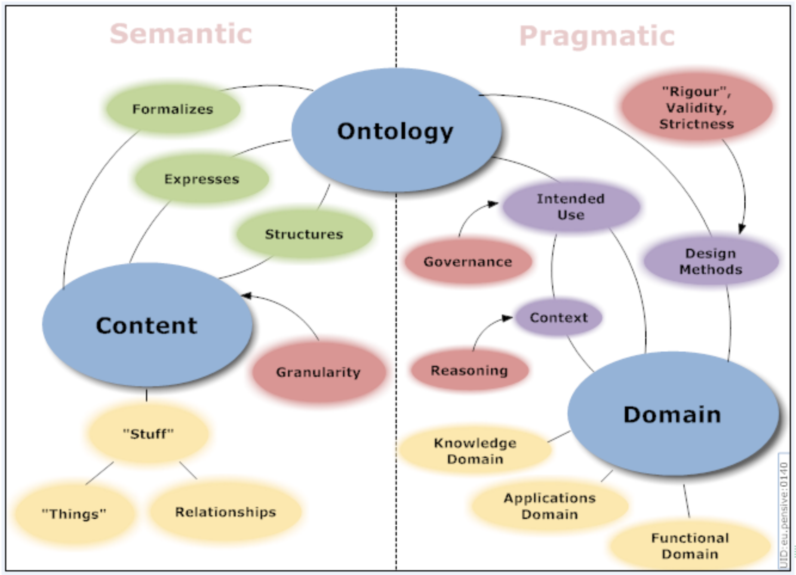
slide from, and more details available in: http://ontolog.cim3.net/file/work/OntologySummit2007/symposium/ OntologyFramework_symposium–Gruninger-Obrst_20070424.ppt

Initial Ontology Dimensions that have Evolved

- Semantic
  - Degree of Formality and Structure
  - Expressiveness of the Knowledge Representation Language
  - Representational Granularity
- Pragmatic
  - Intended Use
  - Role of Automated Reasoning
  - Descriptive vs. Prescriptive
  - Design Methodology
  - Governance

slide from, and more details available in: http://ontolog.cim3.net/file/work/OntologySummit2007/symposium/ OntologyFramework_symposium–Gruninger-Obrst_20070424.ppt

And graphically



Summary

1. Administrivia
2. Introduction
3. Use case: the Semantic Web
4. What is an Ontology?