Knowledge Representation for the Semantic Web

Winter Quarter 2010

Slides 1 – 01/05/2010

Pascal Hitzler
Kno.e.sis Center
Wright State University, Dayton, OH
http://www.knoesis.org/pascal/

Today's Session

1. About me and my expectations for this course
2. About you and your expectations for this course
3. What is *Semantic Web*?
   Why does it need *Knowledge Representation*?
4. Course Project
5. Course contents overview
6. Organizational matters

About me

- 1998 Diplom (Master) in Mathematics, Tübingen, Germany
- 2001 PhD in Mathematics, Cork, Ireland
- 2001-2004 Postdoc in Artificial Intelligence, Dresden, Germany
- 10-12.2003 Research Associate CWRU Cleveland OH
- 2004-2009 Assistant Prof. in Appl. Comp. Sci., Karlsruhe, Germany
- since September 2009: Assistant Prof., Kno.e.sis Center at WSU

- Active Semantic Web researcher – this course is a basic introduction to my current core research area.

- I've done lots of teaching, but never in the US
  I'm happy about all feedback (even anonymous, if you prefer)

My expectations

- **PhD students beyond the Master do not require teaching through courses: If you're pursuing a PhD you should be able to learn all by yourself those things which you need to learn.**

- **Still, learning through courses can be helpful, and can in particular provide shortcuts to things you need.**

- **And it's rarely wrong to acquire broad knowledge.**

- **I expect:**
  That you take responsibility yourself, and decide for yourself how much work you want to put into this course.
  Regretfully, however, I will have to give grades at the end ...

Today's Session

1. About me and my expectations for this course
2. About you and your expectations for this course
3. What is *Semantic Web*?
   Why does it need *Knowledge Representation*?
4. Course Project
5. Course contents overview
6. Organizational matters

Who are you?

- **Which year?**
- **Master or PhD?**
- **Which specialization area?**
- **Why are you here?**
  - **topic relates to my specialization area;**
  - **topic sounds interesting;**
  - **need another course and it doesn't matter which;**
  - **not sure if I'll stay;**
  - **supervisor told me to come;**
  - **a friend dragged me along;**
  - **isn't this the Algebraic Topology class?**
  - **...**

Today's Session

1. About me and my expectations for this course
2. About you and your expectations for this course
3. What is *Semantic Web*?
   Why does it need *Knowledge Representation*?
4. Course Project
5. Course contents overview
6. Organizational matters

The WWW is penetrating our society

- Social contacts (social networking platforms, blogging, ...)
- Economics (buying, selling, advertising, ...)
- Administration (eGovernment)
- Education (eLearning, Web as information system, ...)
- Work life (information gathering and sharing)
- Recreation (games, role play, creativity, ...)

The current Web

- **Immensely successful.**
- **Huge amounts of data.**
- **Syntax standards for transfer of structured data.**
- **Machine-processable, human-readable documents.**

BUT:

- **Content/knowledge cannot be accessed by machines.**
  Meaning (semantics) of transferred data is not accessible.

Limitations of the current Web

- Too much information with too little structure
  and made for human consumption
  - Content search is very simplistic
  - future requires better methods
- Web content is heterogeneous
  in terms of content
  in terms of structure
  in terms of character encoding
  - future requires intelligent information integration
- Humans can derive new (implicit) information from given pieces of information
  but on the current Web we can only deal with syntax
  - requires automated reasoning techniques

Examples

- **Find that landmark article on data integration written by an Indian researcher in the 1990s.**
  [If you manage this without knowing the answer, let me know how you did it.]

- **Are lobsters spiders?**
  [This is getting easier these days, but was impossible a few years ago. It still needs finding and integrating over different websites, as well as some background knowledge.]

- **Which car is called a “duck” in German?**
  [This needs some intelligent integration of content from different websites plus background knowledge.]

Another example

“Identify congress members, who have voted “No” on pro environmental legislation in the past four years, with high-pollution industry in their congressional districts.”

In principle, all the required knowledge is on the Web – most of it even in machine-readable form.

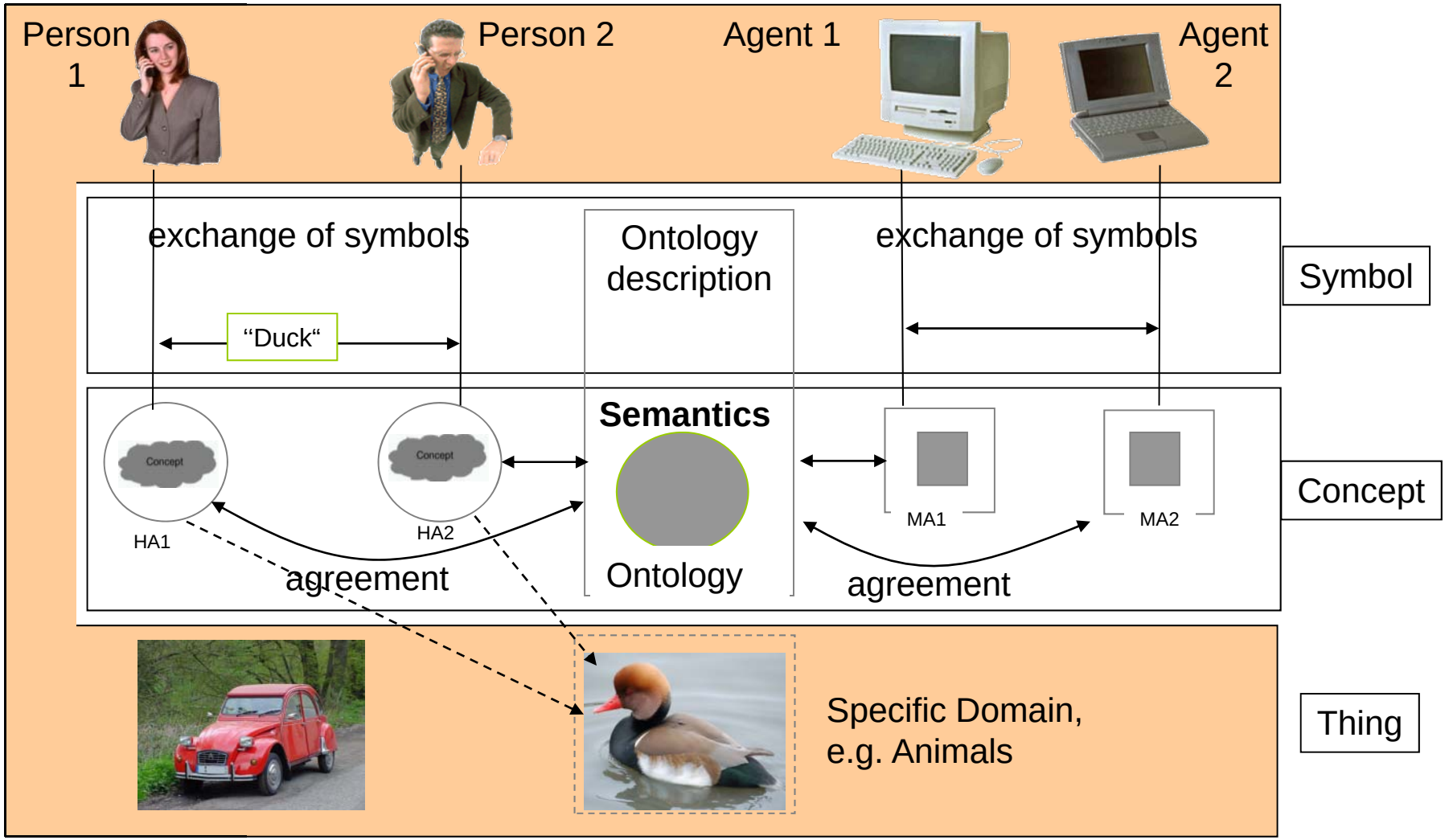
However, without automated processing and reasoning we cannot obtain a useful answer.

Basic ingredients for the Semantic Web

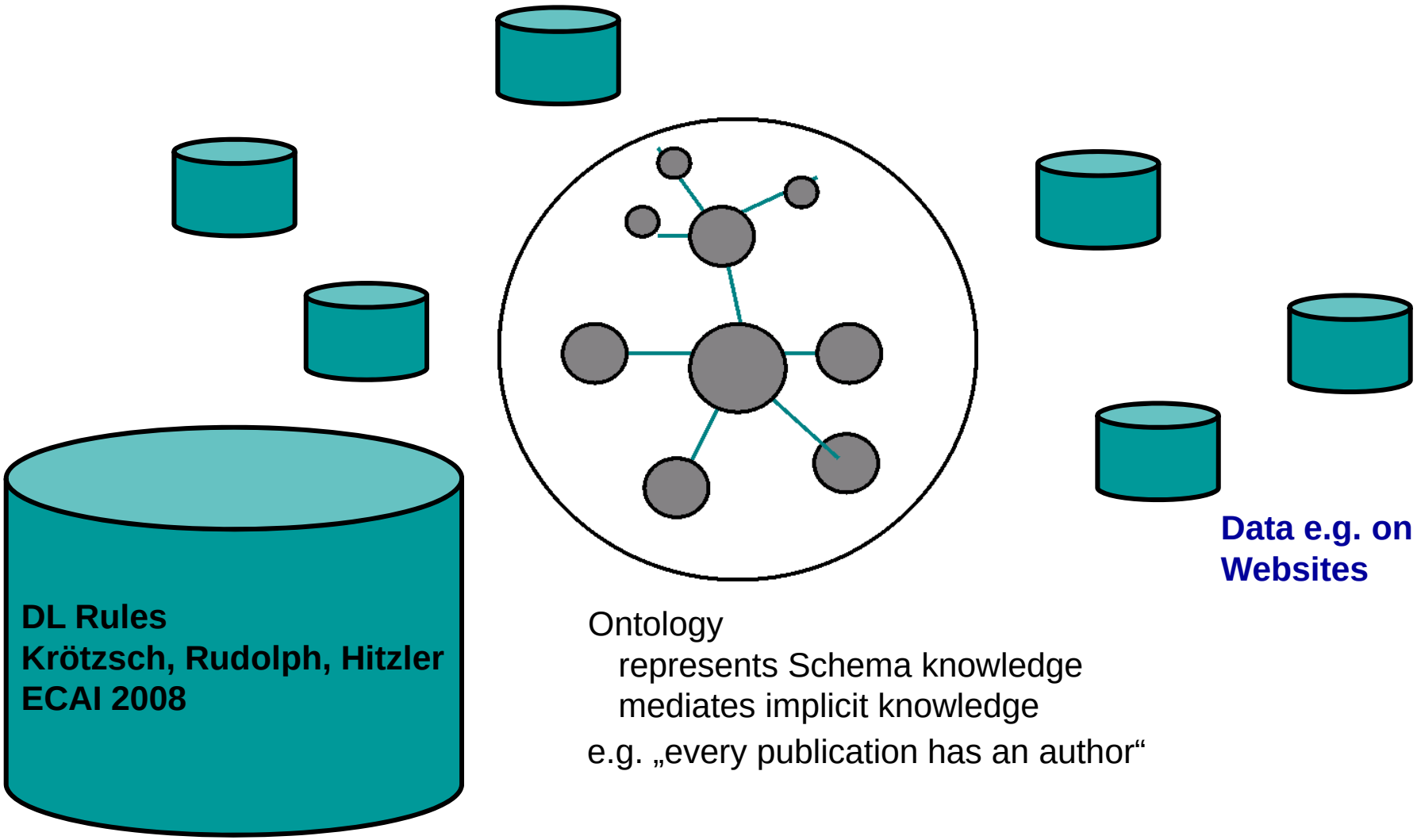
- **Open Standards for describing information on the Web**
- **Methods for obtaining further information from such descriptions**

We'll talk about these matters in this course.

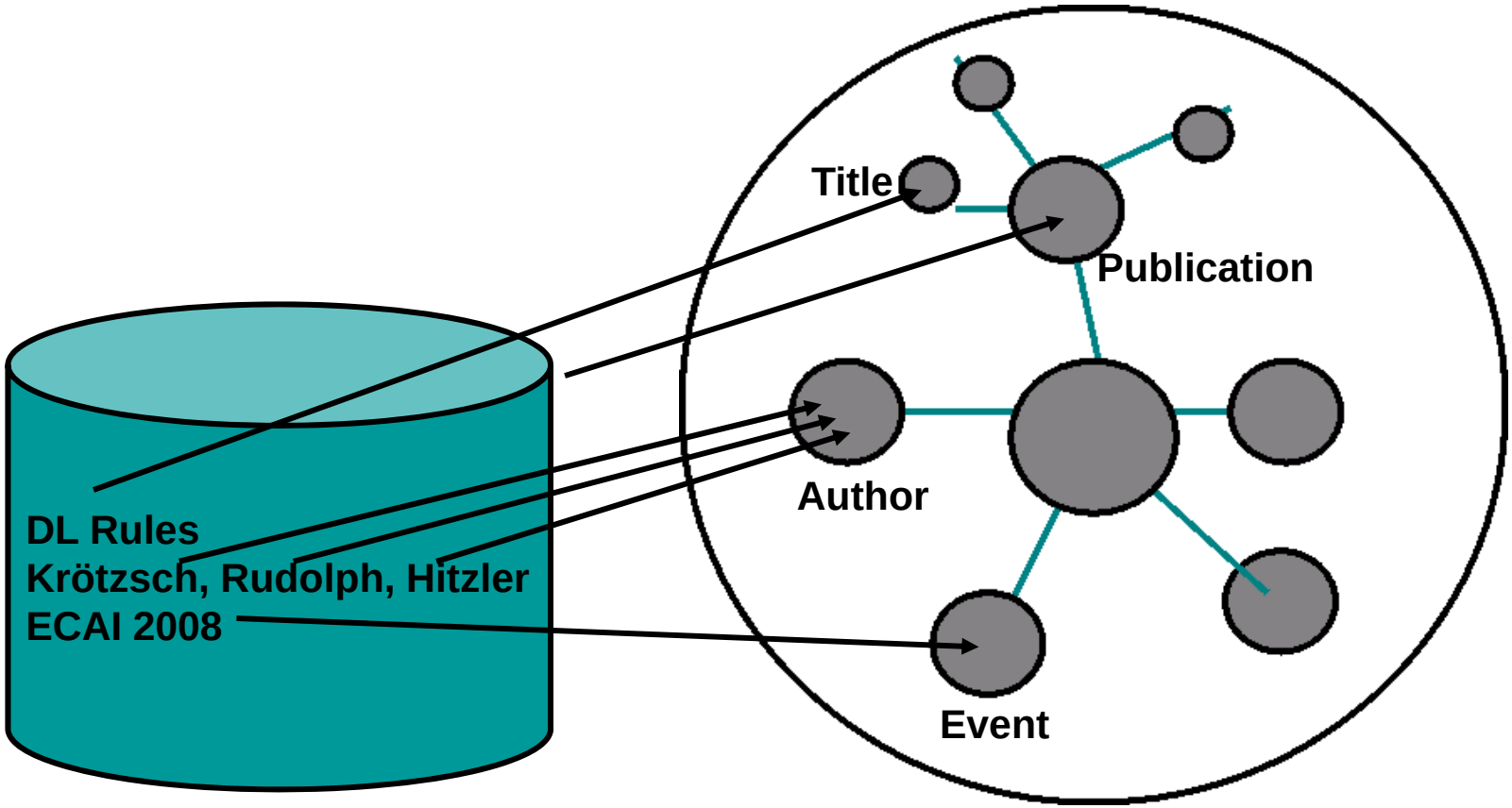
Basic Idea of the Semantic Web



Basic Idea of the Semantic Web

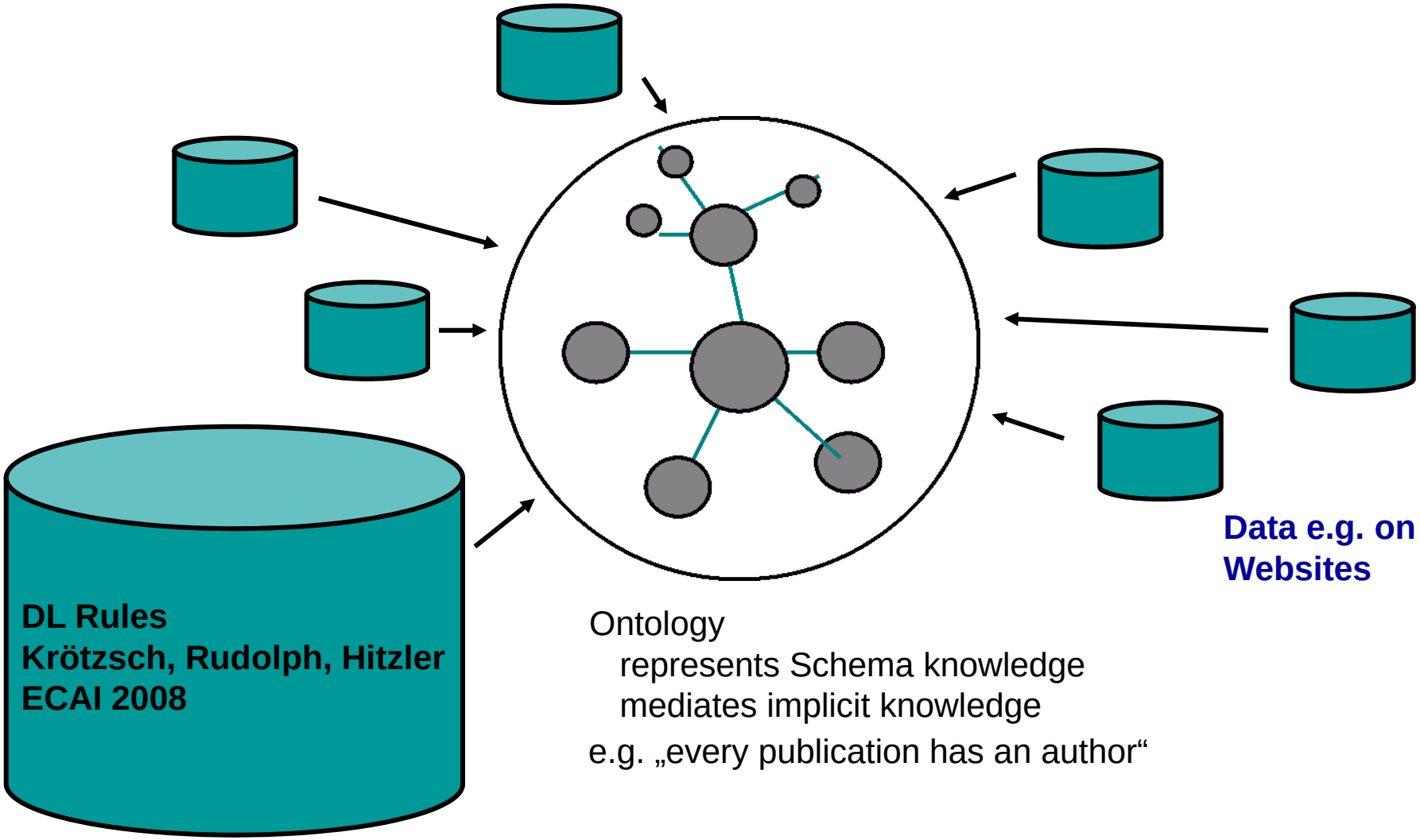


Basic Idea of the Semantic Web



e.g. „every publication has an author“

Basic Idea of the Semantic Web



Ontology
- represents Schema knowledge
- mediates implicit knowledge

e.g. „every publication has an author“

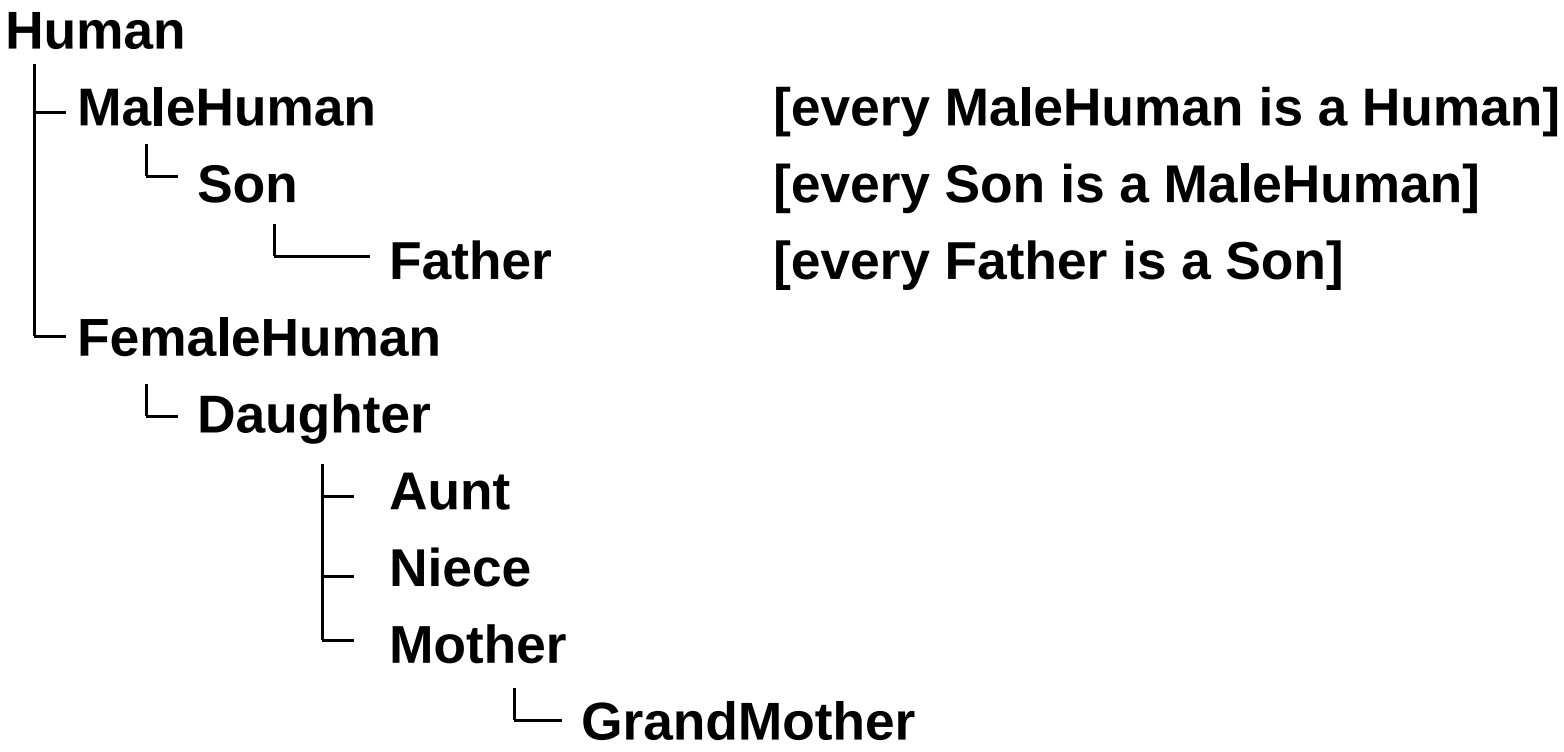
Ontology languages

- **Of central importance for the realisation of Semantic Technologies are suitable representation languages.**
- **Meaning (semantics) provided via logic and deduction algorithms (automated reasoning).**
- **Scalability is a challenge.**

Expressivity Scalability

Ontologies

- **The core of an ontology is usually a *taxonomy*:**
  - **classes of things, arranged in a hierarchy**



Ontologies

- **We will talk more (in fact, a lot) about ontologies**

- **For now, let's focus on taxonomies**

Today's Session

1. About me and my expectations for this course
2. About you and your expectations for this course
3. What is *Semantic Web*?
   Why does it need *Knowledge Representation*?
4. Course Project
5. Course contents overview
6. Organizational matters

Course Project

- Throughout the course, each of you will create an ontology
- We'll do this step by step
- There will also be some interactive and group tasks

- The goal of the project is to learn "hands-on" about ontology modeling

- You'll be given incremental assignments to work on your ontology.

First project assignment

- Select a domain which you want to model (e.g., “family”).
  - use a domain you have good knowledge about
  - use a domain which is accessible for others (e.g. do not choose Algebraic Topology)
  - you will be stuck with this domain until the end of the project!
  - send name of the domain to me by 01/06/2010.
- Select approx. 20 class names for an initial taxonomy.
  - Taxonomy must be at least 3 nodes deep.
- Create the taxonomy and write it up.
  - choose your own representation method – but it must be generally understandable (or give an explanation)
  - send taxonomy to me by 01/10/2010
- Write a few sentences why you chose this domain.
  - send this to me by 01/10/2010

Course Project

- **Deadlines are hard.**
- **The more “difficult” domains are the more interesting ones ;)**
- **You will be able to revise your taxonomy later.**
- **Do this first part without discussion with other students.**
- **Expect that, at some stage, other students will have access to your ontology. I may also discuss examples explicitly in class or on slides. So whatever you write up will be “public.”**
- **Don’t worry (yet!) too much about “correctness” of your modeling.**

Today's Session

1. About me and my expectations for this course
2. About you and your expectations for this course
3. What is *Semantic Web*?
   Why does it need *Knowledge Representation*?
4. Course Project
5. Course contents overview
6. Organizational matters

Course contents overview

- Resource Description Framework (RDF)
  - Syntax
  - Semantics
  - Automated Reasoning
- Web Ontology Language (OWL)
  - Syntax
  - Semantics
  - Automated Reasoning

The rest is not fixed and will depend also on your interests.

Course contents discussion

- What's your knowledge about XML?
- What's your knowledge about predicate logic?

- Possible further topics – we can decide on these later:
  - SPARQL [RDF query language]
  - Querying OWL
  - W3C Rule Interchange Format
  - Rules and Ontologies
  - Ontology Engineering tools
  - Ontology Engineering methods (research papers)
  - ...

- Let me know if there's anything you're particularly interested in.

Today's Session

1. About me and my expectations for this course
2. About you and your expectations for this course
3. What is *Semantic Web*?
   Why does it need *Knowledge Representation*?
4. Course Project
5. Course contents overview
6. Organizational matters

Organizational matters

- **Course Website:**
  http://www.semantic-web-book.org/page/KR4SW-10
  When do you want me to put up the slides?
  What about using WebCT etc?
- **Class meetings:**
  - **Tuesday 2:15pm to 3:55pm in 355 Russ**
  - **Thursday 2:15pm to 3:55pm in 365 Joshi**
  - **Class starts 5th of January 2010 and ends 11th of March 2010**
- **"Office hours:**
  - **Tuesdays 4:15pm to 5:15pm and by appointment.**
    I will also be available after class.
  - **Please use email as main means of communication with me (besides talking with me in or after class).**

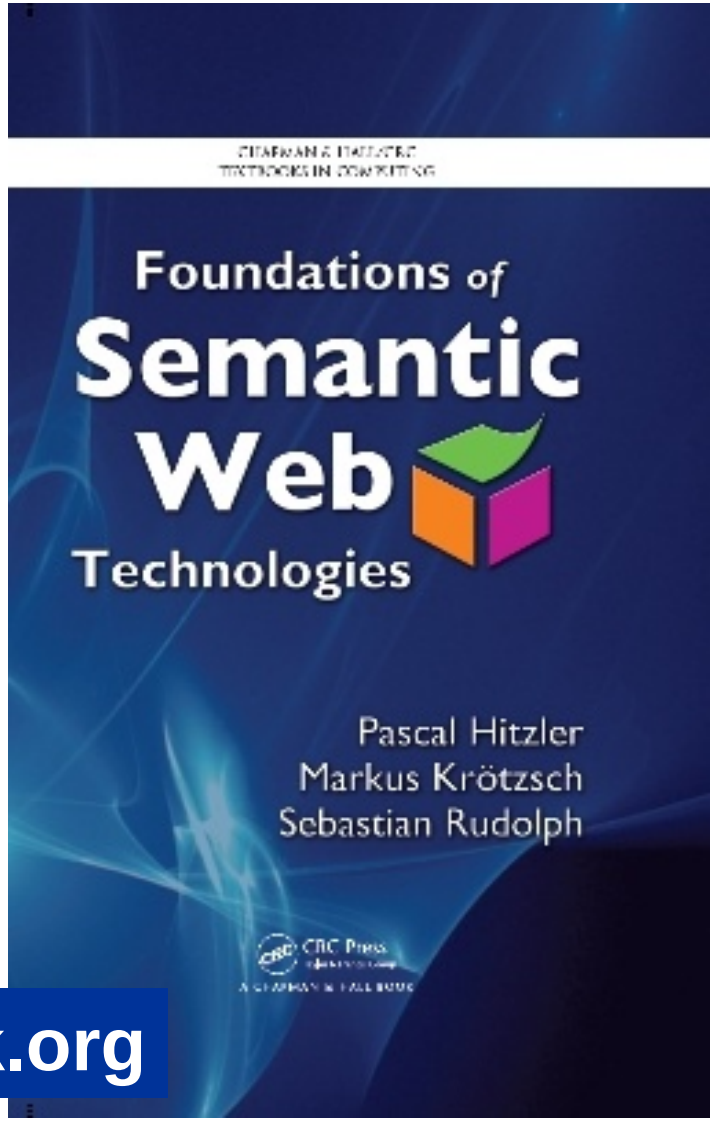
Textbook (required)

Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph

Foundations of Semantic Web Technologies

Chapman & Hall/CRC, 2010

Flyer with special offer is available.



http://www.semantic-web-book.org

Grading

Homework 10%, Project 30%, Class Presentation 30%, Final Exam 30%

- **Homework: Students will take turns in presenting completed homework in class, and grading will be done by evaluating the participation in the tutorial sessions.**
- **Project: The project will be an ontology modeling project which will have several parts and run over the whole quarter. Students are expected to write an experience report on each part.**
- **Class Presentations: The class presentations will be assigned in the first week. Options are presentation of prominent ontology software tools or of original research papers.**
- **Final Exam: The final exam will be oral, i.e. in the form of a short interview (20 minutes), with the examiner asking questions and the student answering.**

Final Exam (oral, 20 minutes)

- **We will probably have to do this in the last week of the lecture period (most likely on 11th + 12th of March).**
- **Are there any conflicts with this?**
- **Any questions about oral exams?**

Class Presentations

- **Options:**
  - **Presentations of Ontology Engineering tools.**
  - **Series of presentations on Ontology Evaluation (Thesis by Denny Vrandecic)**
  - **Presentations of original research papers extending beyond the material covered in this lecture.**

Any volunteers for presenting the Protege ontology modeling system?

Give me feedback on this, and we'll discuss it again next week.

Any further questions or open issues?

Topic next Thursday: TBD :)

First project assignment

- Select a domain which you want to model (e.g., “family”).
  - use a domain you have good knowledge about
  - use a domain which is accessible for others (e.g. do not choose Algebraic Topology)
  - you will be stuck with this domain until the end of the project!
  - send name of the domain to me by 01/06/2010.
- Select approx. 20 class names for an initial taxonomy.
  - Taxonomy must be at least 3 nodes deep.
- Create the taxonomy and write it up.
  - choose your own representation method – but it must be generally understandable (or give an explanation)
  - send taxonomy to me by 01/10/2010
- Write a few sentences why you chose this domain.
  - send this to me by 01/10/2010