Ontoknowledge:
Intelligent Information Brokering in Intranets

Dieter Fensel et al.
Vrije Universiteit Amsterdam

The message of the talk in a nutshell

- The computer was invented as a device for computation.
- Then the PC was detected as a means for games, text processing and power point presentations
- Meanwhile the “computer” becomes a portal to cyberspace

==> The “computer” is in fact an entry point to a world-wide network of information exchange and business transactions.

==> Technology that supports access to unstructured, heterogeneous and distributed information and knowledge sources will become as essential as programming languages were in the 60’s and 70’s.

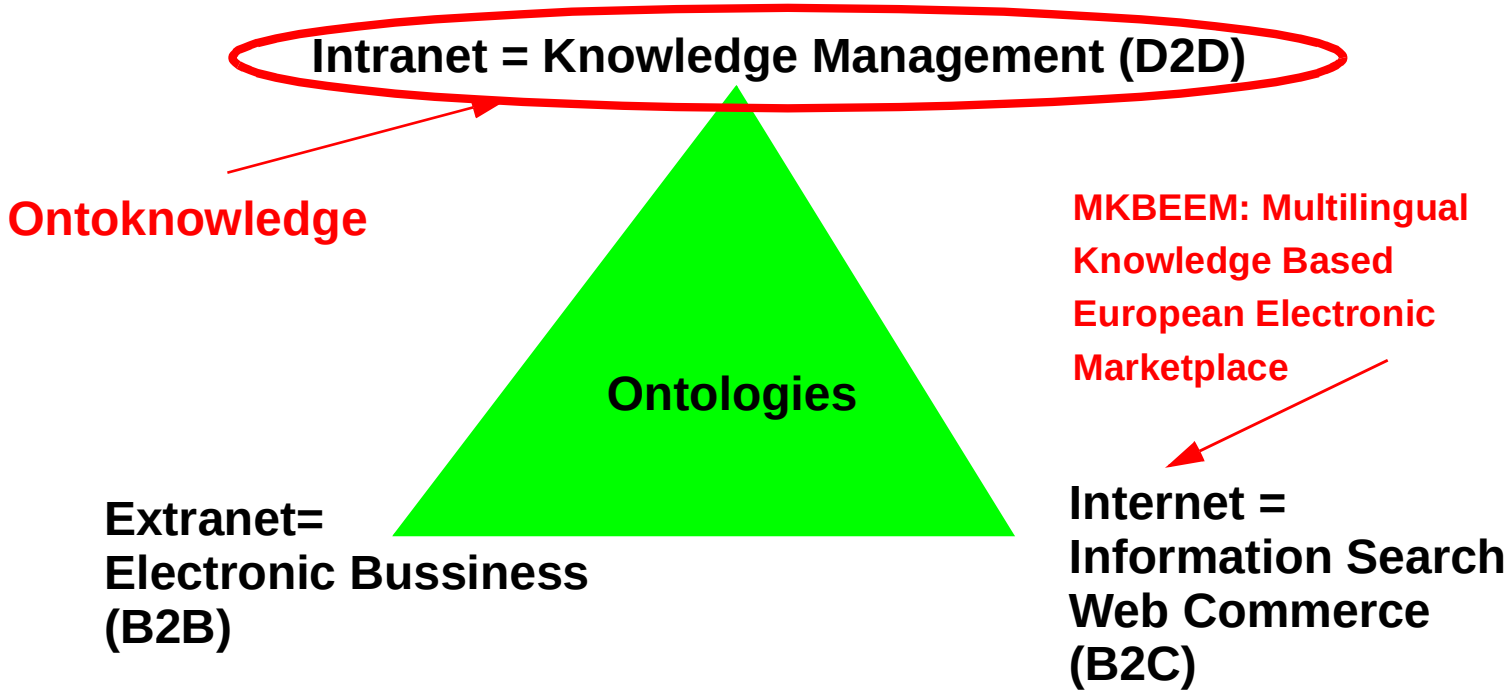
==> At the begin of 2000, we already know the name of this technology. It is called ...

Ontology

Contents

1. The Ontology Triangle
2. Intranet & Internet =
   Information Search & Knowledge Management
3. Ontoknowledge: The whole Picture
4. Ontoknowledge: Milestone I
5. Ontoknowledge: Risk Analysis
6. The vision of a semantic web or knowledge web

1 The Ontology Triangle



==> One technology, three application areas.

Ontologies

- They have been developed in Artificial Intelligence to facilitate *knowledge sharing and reuse*.
- Meanwhile a popular research topic in various communities
  - knowledge engineering
  - natural language processing
  - information systems
  - information retrieval
  - knowledge management
  - electronic commerce
  - knowledge representation, multi-agent systems, ...

Ontologies

- What are ontologies:
  - Ontologies are *consensual* & *formal* specifications of conceptualizations.
  - provide a *shared and common* understanding of a domain that can be communicated across people and application systems

2 Intranet & Internet = Information Search & Knowledge Management

- The competitiveness of companies active in areas with high change rate depends heavily on how they maintain and access their knowledge.
- Main parts of on-line available knowledge of companies can be accessed via Intranets.
- Large Companies have intranets with several million pages.

==> Finding, creating and maintaining information is a rather hard problem in this weakly structured representation media.

Information Search & Knowledge Management

- Knowledge Management should provide methods and tools to deal with these problems.
- Ontoknowledge focus on the technical site of this process.
- It provides methods and tools that support *acquiring*, *maintaining*, and *accessing* knowledge of an organization.
- The case-study driven character of Ontoknowledge ensures that the developed means reflect the actual needs of their prospected users.

2.1 Situation Today

With the large number of on-line documents several document management systems arose. However these systems have severe weaknesses:

- Word matching as search method.
- Information Retrieval instead of Query Answering.
- Document exchange between departments is only possible with severe effort.
- Different views on documents are not supported.

2.2 Near Future

Ontologies will allow structural and semantic definitions of documents providing completely new possibilities:

- Intelligent search instead of keyword matching.
- Query Answering instead of Information Retrieval.
- Document exchange between departments via transformation operators.
- Definition of views on documents.

3 Ontoknowledge: The whole Picture

- Currently the web lacks means for defining structure and semantics of data.
- This is changing with the upcoming new standards XML, XML schema, XSL-T and XQLs, RDF, and RDF Schema. They provide the standardized syntax for defining the semantics of information.
- Ontologies will actually be used to define the semantics of information in a machine-processable manner.

==> **Ontoknowledge** provide a tool environment and a methodology for using these techniques for effective and efficient knowledge management.

3.1 Ontoknowledge: The Goals

The goal of the On-To-Knowledge project is to support efficient and effective knowledge management. It focuses on *acquiring*, *maintaining*, and *accessing* weakly-structured on-line information sources:

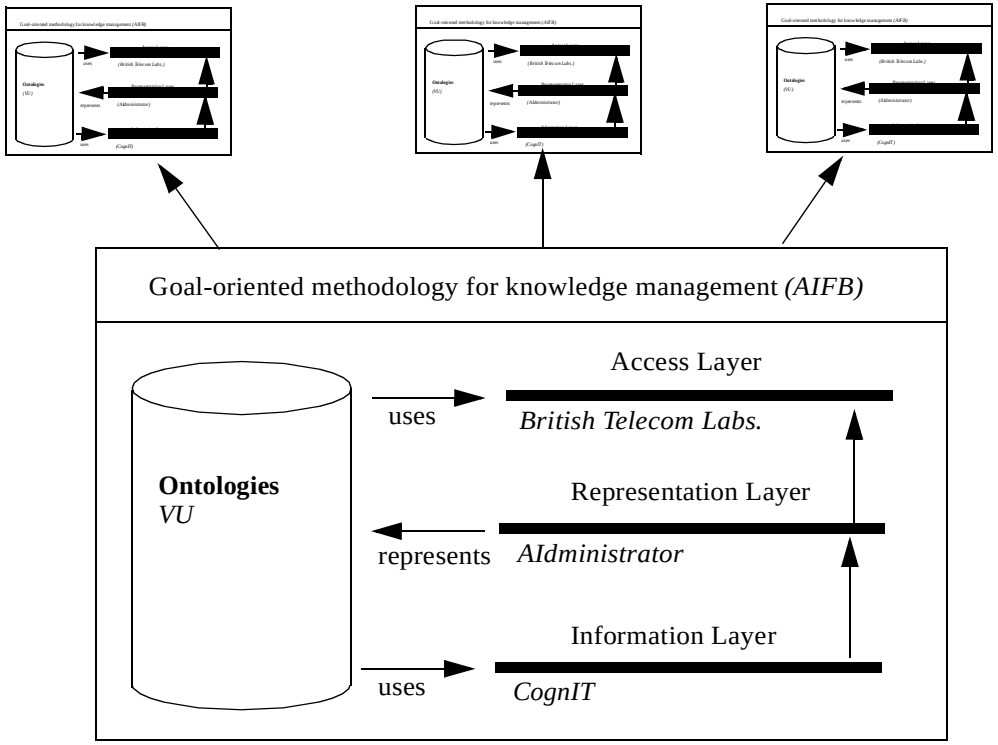
- *Acquiring*: Text mining and extraction techniques are applied to extract semantic information from textual information.
- *Maintaining*: RDF and XML are used for describing syntax and semantics of semi-structured information sources.
- *Accessing*: Push services and agent technology support users in accessing the information.

3.2 Ontoknowledge: The Architecture

Case Study: Org. memory
Swiss Life

Case Study: Call Centre
British Telecom

Case Study: Virtual Enterprise
Enersearch



3.3 Ontoknowledge: The Information Layer (CognIT)

CognIT offers products and services related to both knowledge management and information services. *Corporum* is a state-of-the art tool in information retrieval.

In the course of the project we will extend this tool by:

- semi-automatic information extraction: wrapper building environment for
- automatic information extraction (i.e., automatic annotation of information sources)
- ontology learning

3.4 Ontoknowledge: The Representation Layer (AIdministrator)

AIdministrator products aid in the classification, visualization, navigation, and maintenance of large amounts of semi-structured information.

- We will develop a storage and query system of meta data based on RDF schema.
- This tool is the intermediate layer between the rough information sources and the user interface.

3.5 Ontoknowledge: The Access Layer (BT Labs)

BT is developing a range of products designed for knowledge management: *ProSum*, *ProSearch*, *Jasper*, and the *Information Garden*.

==> Currently, all these tools are based on keyword information retrieval

In Ontoknowledge, we will base these tool on explicit representation of the semantics of the information.

As input for the user interface we provide:

- Annotated information
- Ontologies

3.6 Ontoknowledge: Case Studies (Swiss Life and Enersearch)

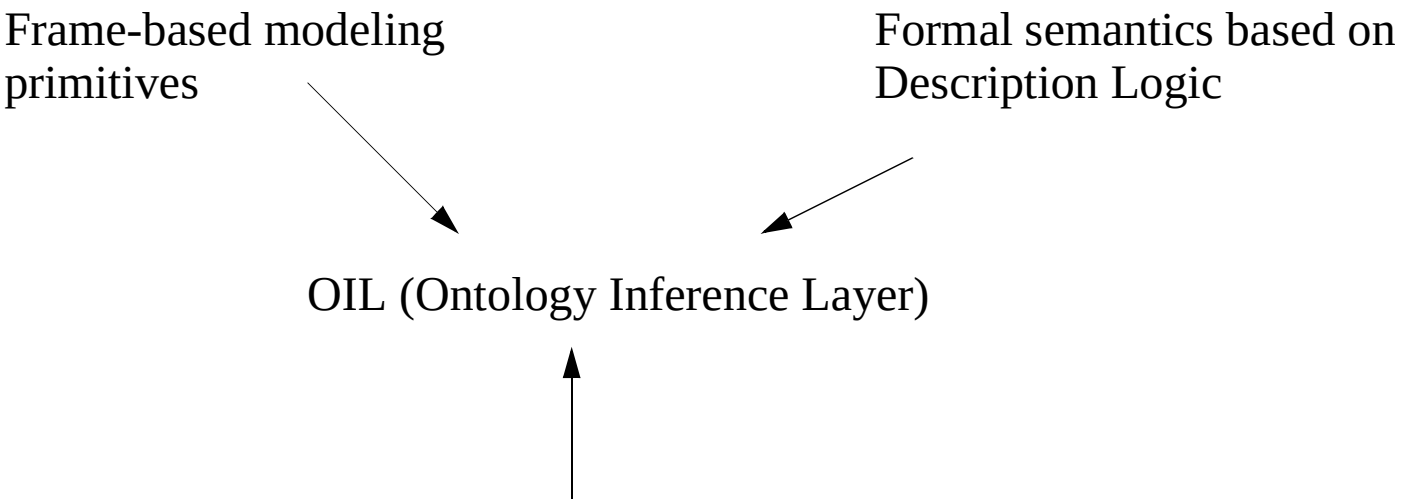
- **Swiss Life**: *Large and weakly structured organizational memory.*
  The project will achieve a smooth integration of weakly structured knowledge with formally represented knowledge and a maintenance tool that makes extension and modification of both, weakly structured and formalized knowledge easy.
- **Enersearch**: *Knowledge management in a virtual enterprise.*
  The aim of this case-study by Enersearch is to evaluate the Ontoknowledge toolkit for increased knowledge transfer between different researchers in the partner's organization, and between researchers and specialist from Enersearch's shareholder companies.

3.7 Ontoknowledge: Methodology (AIFB)

- The methodology provides management guidelines and methods ensuring a successful introduction of knowledge management tools into an enterprise.
- It is an integrated approach balancing the organization and management aspects on the one hand, and the information technology and systems aspects on the other hand.
- It is based on existing methods like CommonKADS and the experiences made in the three case studies in Ontoknowledge.

3.8 Ontoknowledge: The Ontological Backbone (VU)

- In cooperation with many external partners we created a proposal for a web-based ontology representation language: OIL
- The three roots of OIL:



4 Ontoknowledge: Milestone I

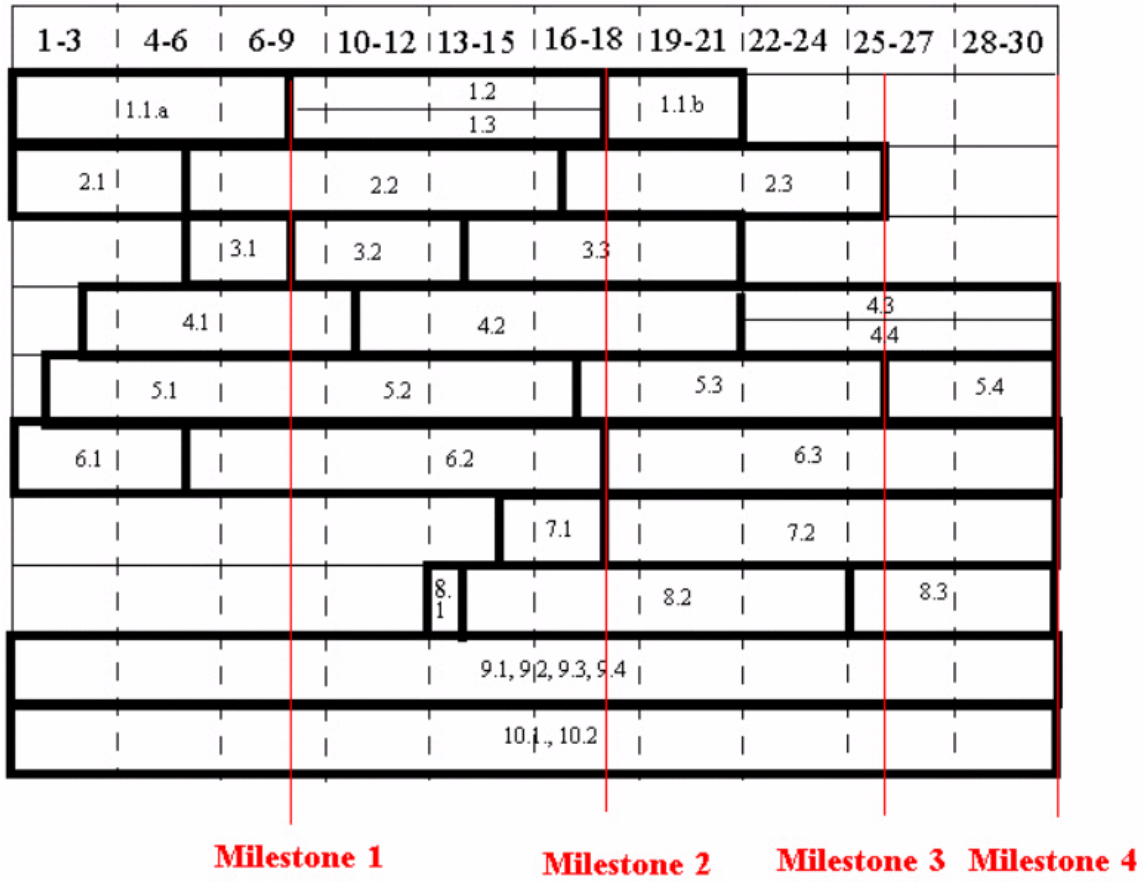
We will discuss:

- The general workplan of Ontoknowledge.
- The goals and steps of milestone I.
- 6-Month Human Resources Allocation per WP and Partner
- The schedule of our meeting today.

4.1 General Workplan

On-To-Knowledge Workpackages			
	Workpackage Title	Main partner	Man Months[1]
WP1	Developing tools for **Ontology Construction and Interchange**	VU	32
WP2	Developing support in **Information extraction**	CognIT	19.5
WP3	Developing an infrastructure for **Information representation**	AIdministrator	24
WP4	Developing various means for **Information access**	BT	44
WP5	**Methodology and Management Guidelines**	AIFB	32
WP6	A **case study** on Organisational memory	Swiss Life	34
WP7	A **case study** on call centres	BT	13
WP8	A **case study** on knowledge management in a virtual enterprise	Enersearch	18.5
WP9	**Information dissimination** and take-up of results	VU	13
WP10	**Project Management** and **Exploitation Management**	VU & BT	22
			252

General Workplan



4.2 Successfully initialize the project (Milestone 1 after 8 months)

- *The **first version of the Ontology language** (D1) is defined in WP-1,*
- the **state of the art report on information extraction** is available (D5) in WP-2,
- the **state of the art report on representation formalisms** is available (D8) in WP-3,
- *the **base line version of the knowledge management methodology** (D15) is developed in WP-5,*
- the **first phase of the case study on Organizational Memory** (D19) has been done in WP-6,
- the ***web site** of the project (D30)*, the **project presentation** (D33) and the ***dissemination and use plan** (D34)* are available in WP-9.

Essential Deliverables: D1, D15, D30, D34

4.3 6-Month Human Resources Allocation per WP and Partner

Partner	1	2	3	4	5	6	7	8	9	10	Total
AIdm	3		4			1				0.4	**8.4**
AIFB	3				3						**6**
BT	1	1	2	2					0.3	1	**7.3**
CognIT		5									**5**
Enersearch											
Swiss Life						7.3				0.2	**7.5**
VU	3.6				0.6				3	3.5	**10.7**
Total	**10.6**	**6**	**6**	**2**	**3.6**	**8.3**			**3.3**	**5.1**	**44.9**

4.4 Schedule of the Project

- No task has any significant delay.
- In fact, we gave higher speed to some tasks as it was required by the contract (for example, the query engine).

4.5 Schedule of today = main deliverables

- 09:30 - 10:30

  Ontoknowledge: Intelligent Information Brokering in Intranets (Del 33)

- 10:30 - 11:15 Industrial Case Studies of Ontoknowledge (Del 19)

- 11:30 - 12:15 OIL: Ontology Inference and Interchange (Del 1)

- 14:00 - 14:45 Knowledge Management Methodology (Del 15)

- 14:45 - 15:15 Ontoknowledge: Dissemination and Use (Del 30, Del 34)

5 Ontoknowledge: Risk Analysis

- Will our results be of any use?
- Will our results be important?
- Do we work together properly to guarantee that we will achieve some significant results?
- Do we already have some results?

5.1 Will our results be of any use?

Yes, very likely!

- Ontoknowledge is driven by three major case studies ensuring the method and tool development is in close interaction with users and their actual needs.
- The case study of Swiss Life already triggered many requirements for the user interface.

5.2 Will our results be important?

Our approach is based on machine-processable semantics of information as means for knowledge management.

- *Machine-processable semantics of information sources* or the so-called **semantic web** becomes currently a major subject for research and industry. DARPA already decided to spent 70 Million Dollar on this (www.daml.org) and the EC sets up a new action line in key action III with around 20 Million Euro.
- *Knowledge Management* is a “hot topic” for many companies and IT departments anyway.
- Our language is OIL is used by a large and world-wide community.

5.3 Do we work together properly to guarantee that we will achieve some significant results?

Yes, we do!

- We had three plenary meetings, many meetings with a part of the consortium, and many bilateral meeting and cooperations (especially between the company partners).
- Main bottleneck for this cooperation is the limited amount of travelling money. The budget will not be enough to cover the travelling costs!

5.4 Do we already have some results?

Yes, we do! However, given the early stage of the project this is mainly paper work.

- D. Fensel, I. Horrocks, F. Van Harmelen, S. Decker, M. Erdmann, and M. Klein: OIL in a nutshell. In Knowledge Acquisition, Modeling, and Management, Proceedings of the European Knowledge Acquisition Conference (EKAW-2000), R. Dieng et al. (eds.), Lecture Notes in Artificial Intelligence, LNAI, Springer-Verlag, October 2000.

- M. Klein, D. Fensel, F. van Harmelen, and I. Horrocks: The relation between ontologies and schema-languages: Translating OIL-specifications in XML-Schema. In Proceedings of the Workshop on Applications of Ontologies and Problem-solving Methods, 14th European Conference on Artificial Intelligence ECAI'00, Berlin, Germany August 20-25, 2000.

Do we already have some results?

- D. Fensel, M. Crubezy, F. van Harmelen, and I. Horrocks: OIL & UPML: A Unifying Framework for the Knowledge Web. In Proceedings of the Workshop on Applications of Ontologies and Problem-solving Methods, 14th European Conference on Artificial Intelligence ECAI'00, Berlin, Germany August 20-25, 2000.
  - D. Fensel, F. van Harmelen, H. Akkermans, M. Klein, J. Broekstra, C. Fluyt, J. van der Meer, H.-P. Schnurr, R. Studer, J. Davies, J. Hughes, U. Krohn, R. Engels, B. Bremdahl, F. Ygge, U. Reimer, and I. Horrocks: OnToKnowledge: Ontology-based Tools for Knowledge Management. In Proceedings of the eBusiness and eWork 2000 (EMMSEC 2000) Conference, Madrid, Spain, October 2000.
- S. Decker, D. Fensel, F. van Harmelen, I. Horrocks, S. Melnik, M. Klein, and J. Broekstra: Knowledge Representation on the Web. In Proceedings of the 2000 International Workshop on Description Logics (DL2000), Aachen, Germany, August 2000.

Do we already have some results?

- S. Decker, F. van Harmelen, D. Fensel, I. Horrocks, S. Melnik, M. Erdmann, M. Klein, and J. Broekstra: Knowledge Representation on the Web, to appear in IEEE Internet Computing.
- Maedche, A., Schnurr, H.-P., Staab, S. and Studer, R.: Representation Language-Neutral Modeling of Ontologies. In J. Ebert et al. (eds.), Modelle und Modellierungssprachen in Informatik und Wirtschaftsinformatik, Foelbach Verlag, Koblenz, 2000.

- Staab, S. and Maedche, A.: Ontology Engineering beyond the Modeling of Concepts and Relations. to appear: A. Gomez-Perez. Proceedings of he ECAI'2000 Workshop on Applications of Ontologies and Problem-Solving Methods, Berlin, 2000.
- D. Fensel: The role of Ontologies in Information Interchange. In Proceedings of the 2nd International Scientific and Practical Conference on Programming UkrPROG'2000, Ukraine, Kiev, May 2000.

Do we already have some results?

- Dieter Fensel: Relating Ontology Languages and Web Standards. In J. Ebert et al. (eds.), Modelle und Modellierungssprachen in Informatik und Wirtschaftsinformatik, Foelbach Verlag, Koblenz, 2000.
- Besides publications, we have set up a large international initiative on ontology language standardization with OIL.
  http://www.ontoknowledge.org/oil
- VU is leading an initiative on a thematic network called **OntoWeb** dealing with these issues (http://www.cs.vu.nl/~ying/OntoWeb).

6 The vision of a semantic web or knowledge web

- The World Wide Web is a big and impressive success story, both in terms
  - of the amount of available information and
  - of the growth rate of human users.
- It starts to penetrate most areas of our daily life and business.
- This success is based on its simplicity.

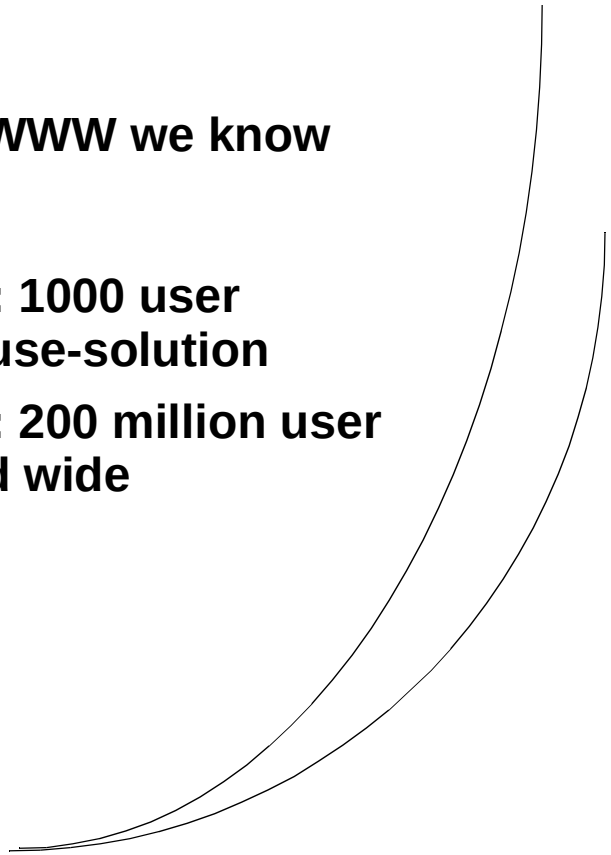
==> The restrictiveness of HTTP and (early) HTML allowed software developers, information provider, and users to make easy access of to new media helping it to reach a critical mass.

The vision of a semantic web or knowledge web

- However, this simplicity may hamper the further development of the Web.
- Or in other words:

What we see currently is the very first version of the web and the next version will probably even more bigger and much more powerful compared to what we have now.

The vision of a semantic web or knowledge web



The WWW we know

1990: 1000 user inhouse-solution

2000: 200 million user world wide

The WWW we will encounter soon

2005: > 1 billion *human* user world & *device* wide

and how many billions of non-human users?

The vision of a semantic web or knowledge web

Tim Berners-Lee has a vision of a *semantic web* which

- has machine-understandable semantics of information, and
- trillions of small specialized reasoning services that provide support in automated task achievement based on the accessible information.

==> This gives a completely new perspective for the knowledge acquisition, knowledge engineering, and knowledge representation communities.

The vision of a semantic web or knowledge web

- Twenty years ago, researchers in AI coined the slogan:

  "***knowledge is power***".

- Quickly two communities arose:
  - knowledge acquisition/engineering deals with the bottleneck of acquiring and modeling knowledge (human-oriented problem).
  - knowledge representation deals with the bottleneck on representing knowledge and reason about (computer-oriented problem).

- However, the results of both communities never really hit the nail.
  - Knowledge acquisition was too costly
  - Developed systems where mainly isolated, brittle, and small solutions for minor problems.

The vision of a semantic web or knowledge web

- Then Tim came around and made a simple trick leading to 100 millions of knowledge acquisitioners (working nearly for free).
- The transformation of the web to the knowledge web suddenly puts KA and KR in the centre of an extremely interesting and powerful topic.
- Given the amount of the knowledge in the web that we already have achieved, this knowledge web will be an extremely knowledgeable, useful, and powerful device.

The vision of a semantic web or knowledge web

- Imagine a web that contains large bodies of the overall human knowledge and trillions of specialized reasoning services that make use of it.
- Compared to the potential of the knowledge web the original AI visions look like a small and old-fashioned idea of the 19th century.
- Darpa already decided to spent 80 million dollar on research for the knowledge web.
- Ontoknowledge is in the centre of this knowledge web providing intelligent access to information sources.