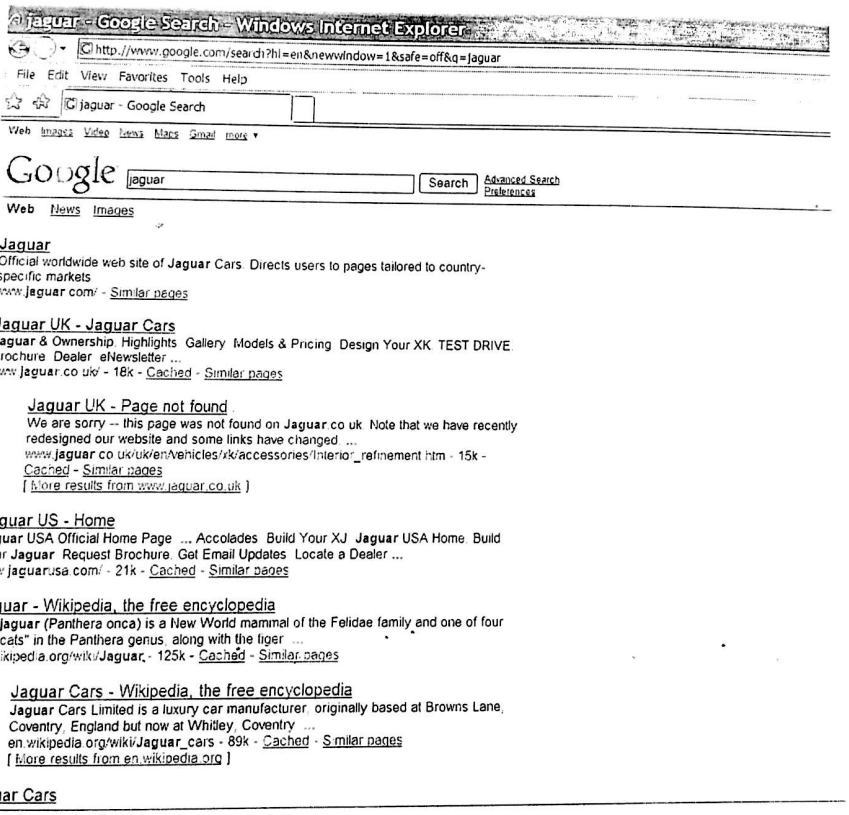
The Web

- What are the main component of the Web?
  - HTTP (how to transfer data)
    - `GET /index.html`
  - URI (how to address data)
    - `http://www.cs.um.edu.mt/....`
  - HTML (how to mark up data for human reader)
    - `<html><head><title>.....`

The core problem of the Web

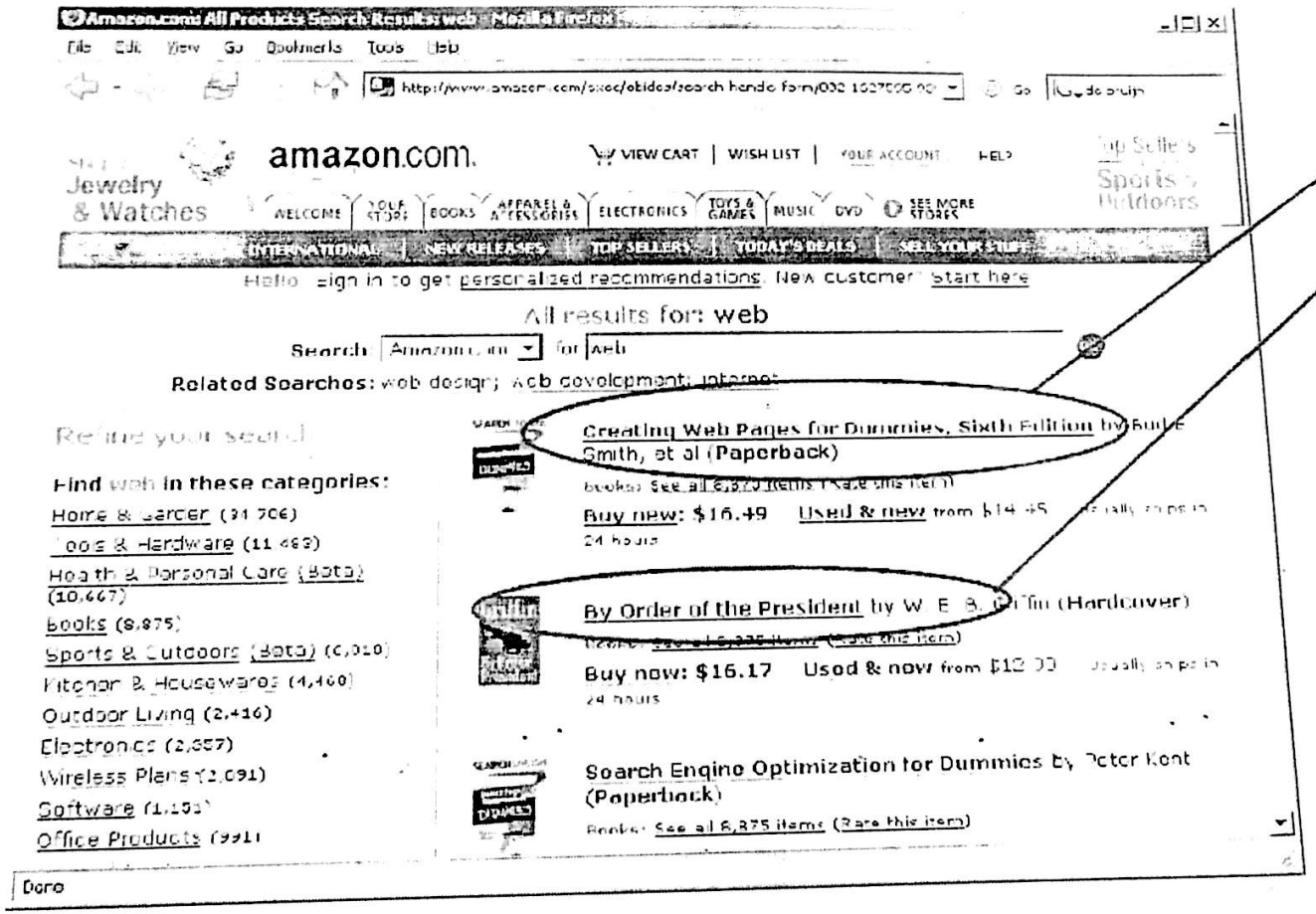
- **Information Overload** which leads to problems when
  - Retrieving documents
  - Extracting relevant data from retrieved documents
  - Combining information from different sources to achieve a particular goal

Retrieve a document

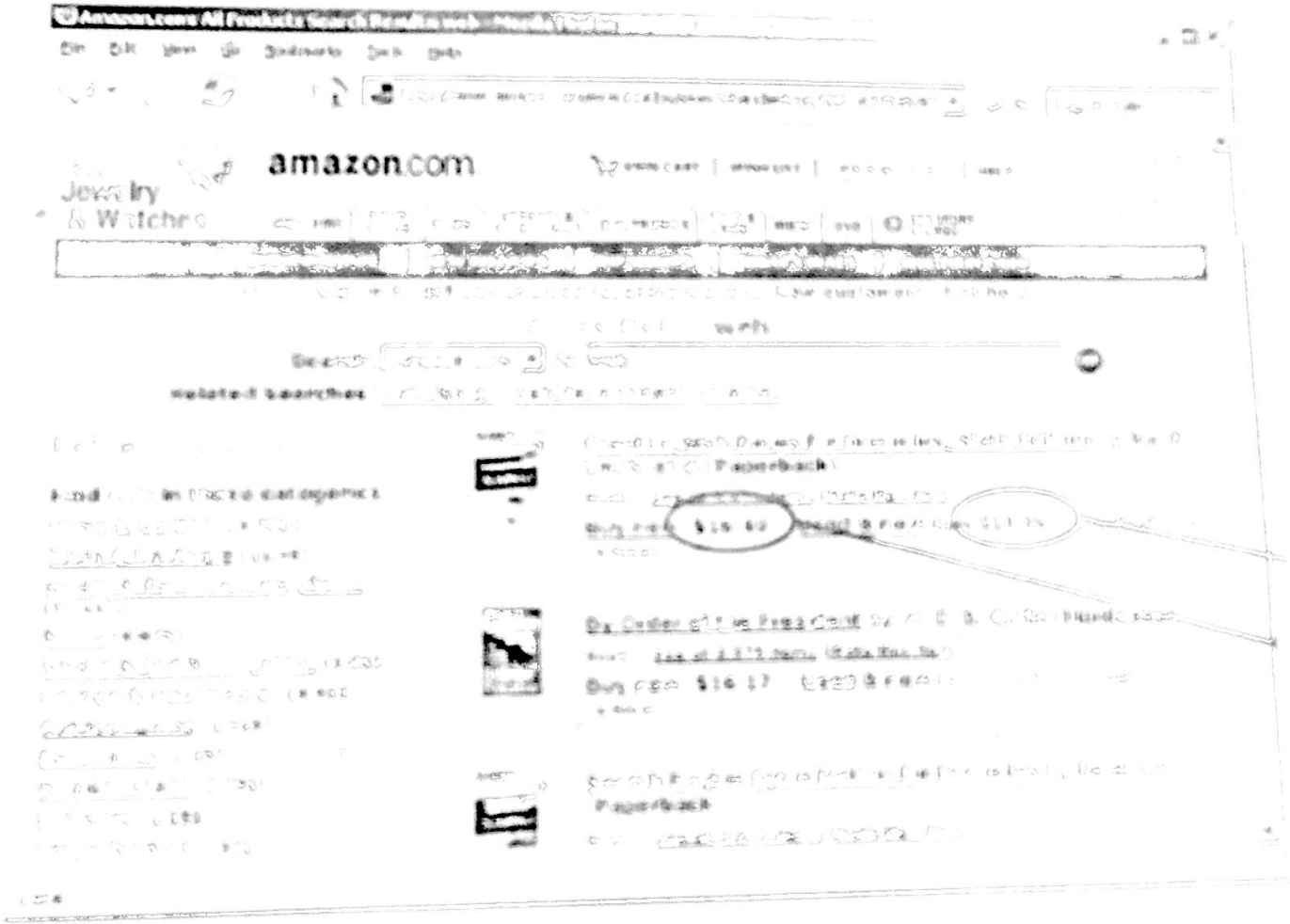


- **Querying for "jaguar" returns various types of results:**
  - Cars
  - Feline
  - Operating system
  - Who knows what else

Extracting information



Extracting information



Aggregating information

Find me the cheapest price for the book "Semantic Web Primer"

SEARCH RESULTS

We found 1 item for keywords: **semantic web primer**.

Sorted by: Top Matches GO

1. A Semantic Web Primer
Grigoris Antoniou, Frank Van Harmelen

Format: **Textbook Hardcover**
Pub. Date: July 2004

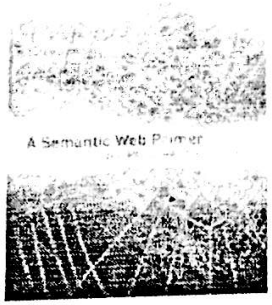
FROM B&N

Online Price: $42.00
Members Pay: **$33.60**

Usually ships within 24 hours

Used Copies Available from our Authorized Sellers

Join Amazon Prime and ship Two-Day for free and Overnight for $3.99. Already a memb



See larger image
Share your own customer images
Publisher: learn how customers can search inside this book.

A Semantic Web Primer (Cooperative Information Systems) (Hardcover)
by Grigoris Antoniou (Author), Frank van Harmelen (Author)
(13 customer reviews)

List Price: $42.00
Price: $31.45 & this item ships for **FREE with Super Saver Shipping**. Details
You Save: $10.55 (25%)

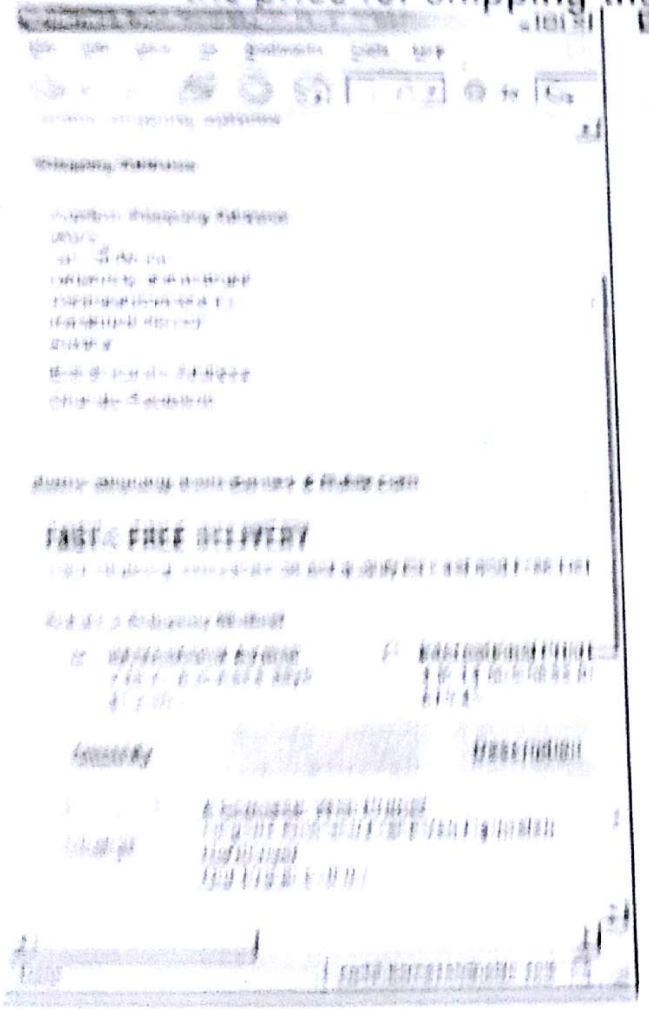
Availability: In Stock. Ships from and sold by **Amazon.com**. Gift-wrap available.

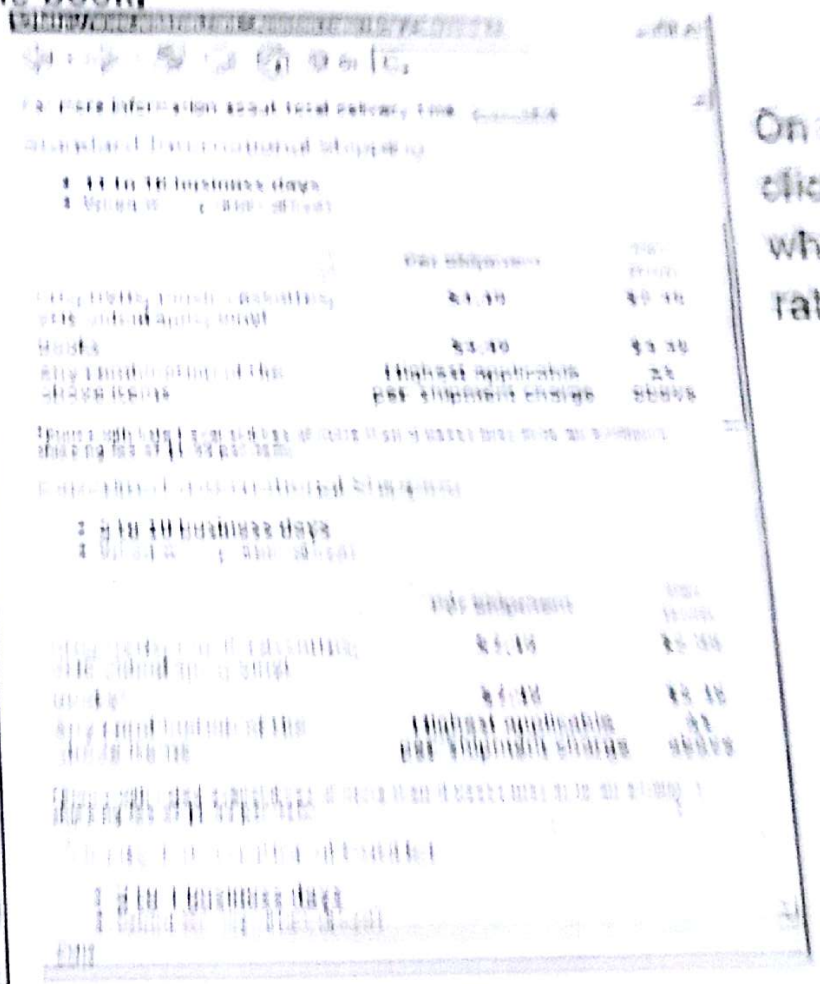
Want it delivered Tuesday, September 25? Order it in the next [illegible] and choose **One-Day**

16 used & new available from $27.30

Aggregating information

I want the cheapest copy of "A Semantic Web Primer"; taking into account the price for shipping the book.



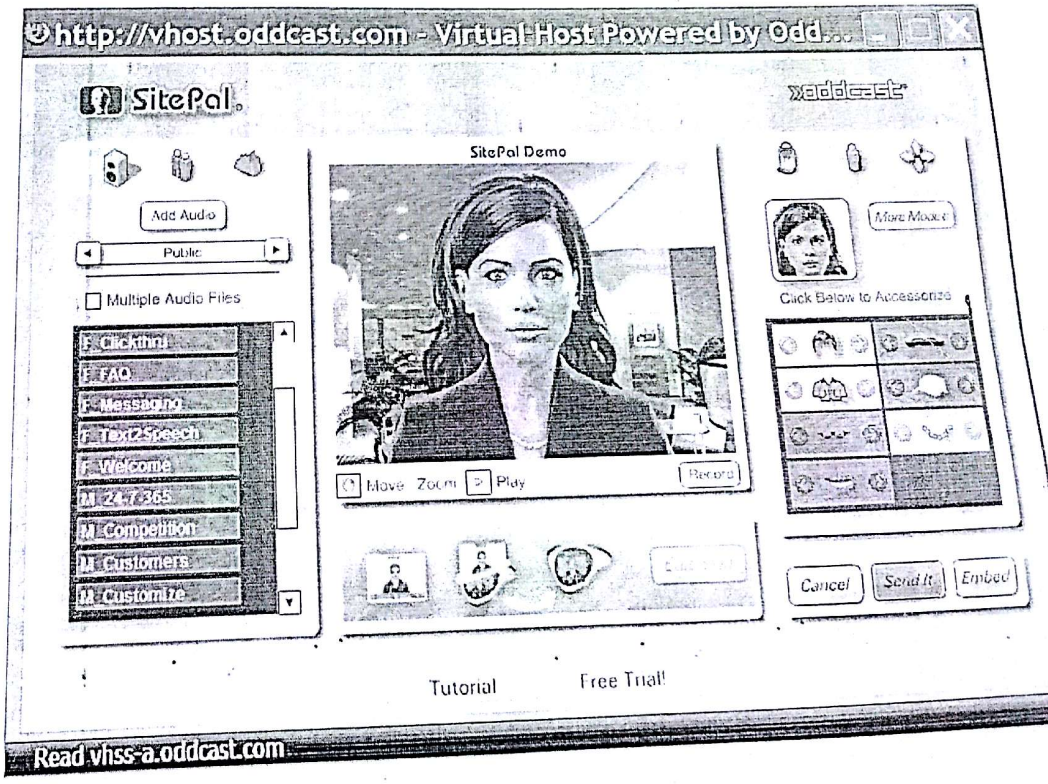


On average 10 clicks to find out what the shipping rate is!

Personal Software Agents

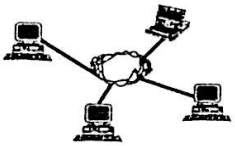
Let a personal assistant handle all the web related tasks. Cool!!

However....



What is Internet Technology?

- What is an internet?
  - Network of networks
- What is *the* Internet?
  - A global internet based on the IP protocol
- To what does "Internet technology" refer? Protocols
  - Architecture
  - Services



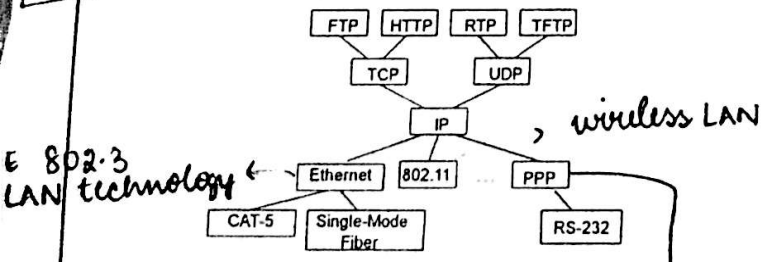
Service-wise (applications)

- Electronic mail
- Remote terminal
- File transfer
- Newsgroups
- File sharing
- Resource distribution
- World Wide Web
- Video conferencing
- Games

Internet "Hourglass" Architecture, defined by the Internet Engineering Task Force (IETF).

Protocols

- **Protocol**
  - *Architecture → Service*
  - *Rules of communication*



, wireless LAN

E 802.3 LAN technology

Standards Making Organizations

ISO = International Standards Organization

ITU = International Telecommunication Union (formerly CCITT)

ANSI = American National Standards Institute

IEEE = Institute of Electrical and Electronic Engineers

IETF = Internet Engineering Task Force

ATM Forum = ATM standards-making body

...and many more

Point-to-Point protocol used to establish direct connection b/w 2 n/wing nodes.

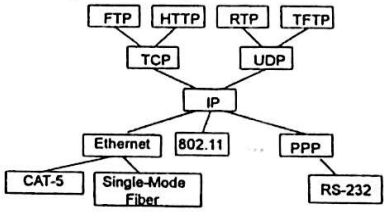
Many different kinds of applical & higher level protocols

Why So Many Standards Organizations?

- Multiple technologies
- Different areas of emphasis and history
  - Telecommunications/telephones
    - ITU,ISO,ATM
  - Local area networking/computers
    - IETF, IEEE
  - System area networks/storage
    - ANSI

Internet "Hourglass" Architecture

- Defined by Internet Engineering Task Force (IETF)
- "Hourglass" Design



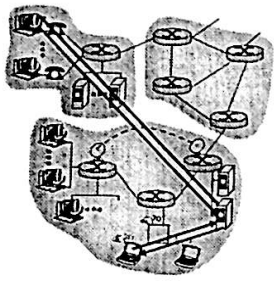
Internet Design Principles

- Scale
  - Protocols should work in networks of all sizes and distances
- Incremental deployment
  - New protocols need to be deployed gradually
- Heterogeneity
  - Different technologies, autonomous organizations
- End-to-end argument
  - Some functions can be correctly implemented at the end hosts; the network should not provided these.

Application architectures

- Client-server
- Peer-to-peer (P2P)
- Hybrid of client-server and P2P

Client-server architecture

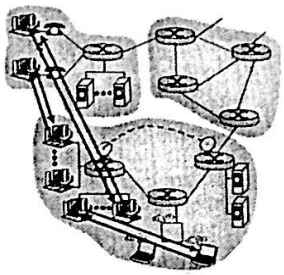


server:
- always-on host
- permanent IP address
- server farms for scaling

clients:
- communicate with server
- may be intermittently connected
- may have dynamic IP addresses
- do not communicate directly with each other

Pure P2P architecture

- not always-on server
- arbitrary end systems directly communicate
- peers are intermittently connected and change IP addresses
- example: Gnutella



Highly scalable but difficult to manage

↳ File sharing mainly

Hybrid of client-server and P2P

Skype
- Internet telephony app
- Finding address of remote party: centralized server(s)
- Client-client connection is direct (not through server)

Instant messaging
- Chatting between two users is P2P
- Presence detection/location centralized:
  - User registers its IP address with central server when it comes online
  - User contacts central server to find IP addresses of buddies

Introduction to the Semantic Web

Payam Barnaghi

The Semantic Web

"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 co-operation.**"

[Berners-Lee *et al*, 2001]

Today's Web

- Currently most of the Web content is suitable for human use.
- Typical uses of the Web today are information seeking, publishing, and using, searching for people and products, shopping, reviewing catalogues, etc.
- Dynamic pages generated based on information from databases but without original information structure found in databases.

Main component of the web

↓ HTTP ↓ URL ↓ HTML

Limitations of the Web Search today

- The Web search results are high recall, low precision.
- Results are highly sensitive to vocabulary.
- Results are single Web pages.
- Most of the publishing contents are not structured to allow logical reasoning and query answering.

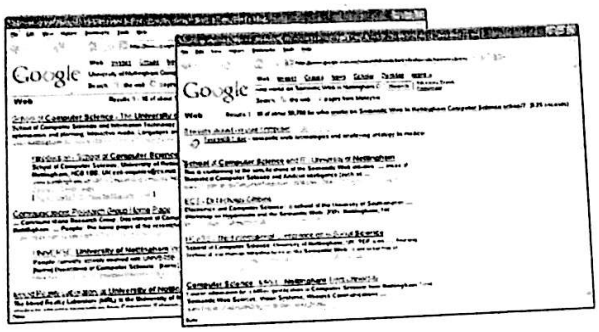
Core problem of the web

INFORMATION OVERLOA

→ extracting relevant data fr retrieved documents

→ combining info. from

Today's Web



5

What is a Web of Data?

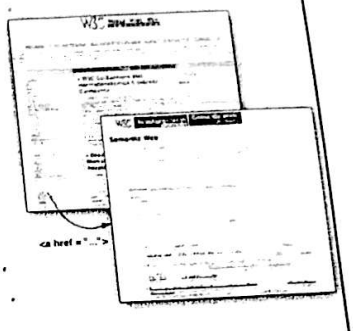
Thinking back a bit... 1994

HTML and URIs

Markup language and means for connecting resources

Below the file level

Stopped at the text level



[Miller 04]

6

What is a Web of Data? (continued)

Now

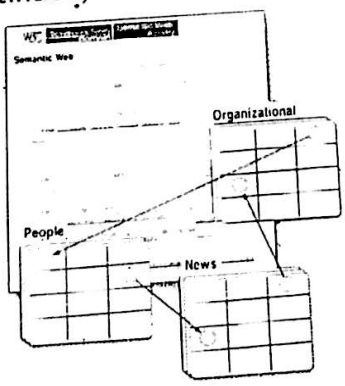
XML, RDF, OWL and URIs

Markup language and means for connecting resources

Below the file level

Below the text level

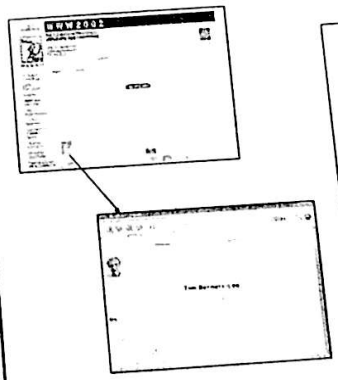
At the data level

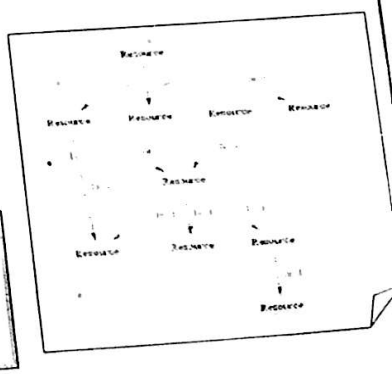


[Miller 04]

7

The Syntactic Web





[Hendler & Miller

8

What is the Problem?

- Consider a typical web page:

- Markup consists of:
  - rendering information (e.g., font size and colour)
  - Hyper-links to related content
- Semantic content is accessible to humans but not (easily) to computers...

[Davies, 03]

i.e. the Syntactic Web is...

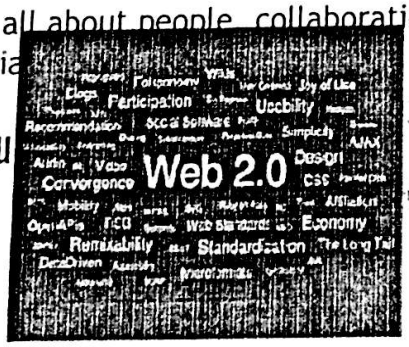
- A place where
  - computers do the presentation (easy) and
  - people do the linking and interpreting (hard).
- *Why not get computers to do more of the hard work?*

[Goble, 03]

Web 2

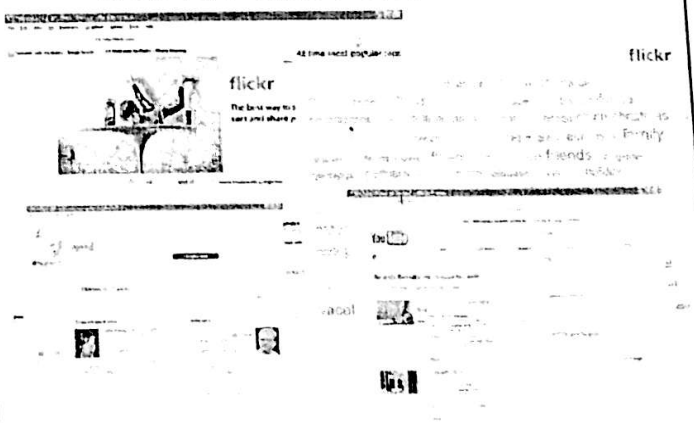
- It is all about people, collaboration, media

You



[The mind-map pictured above constructed by Markus Angermeier, source Wikipedia]

Web 2.0 and Folksonomies



[http://flickr.com/photos/tags/]

Machine-accessible Content

- The main obstacle to provide better support to Web users is that, at present , the meaning of Web content is not machine accessible.
- Although there are tools to retrieve texts, but when it comes to *interpreting* sentence and extracting useful information for the user, the capabilities of current software are still very limited.

Distinguishing the meaning

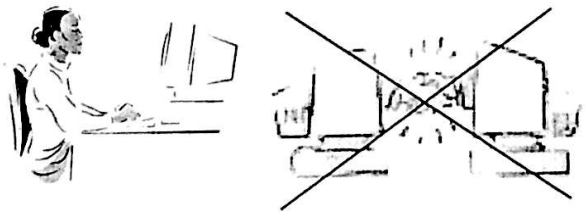
- It is simply difficult for machines to distinguish the meaning of:

  I am a philosopher.

  from

  I am a philosopher, you may think. Well,...

...Limitations of the Web today



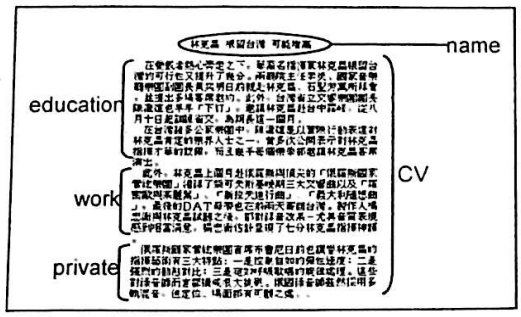
The Web activities are mostly focus on Machine-to-Human, and Machine-to-Machine activities are not particularly well supported by software tools.

[Davies, 01]

How Can the Current Situation be Improved?

- An alternative approach is to represent Web content in a form that is more easily machine-accessible and to use intelligent techniques to take advantage of these presentations.

Machine Accessible Meaning



[Davies, 03]

17

XML

- User definable and domain specific markup

HTML:

Internet and World Wide Web

Code: G52IWW
Students: Undergraduate

XML:

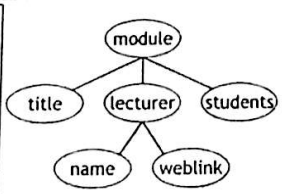
Internet and World Wide Web
G52IWW
Undergraduate

18

XML: Document = labeled tree

- *node = label + contents*

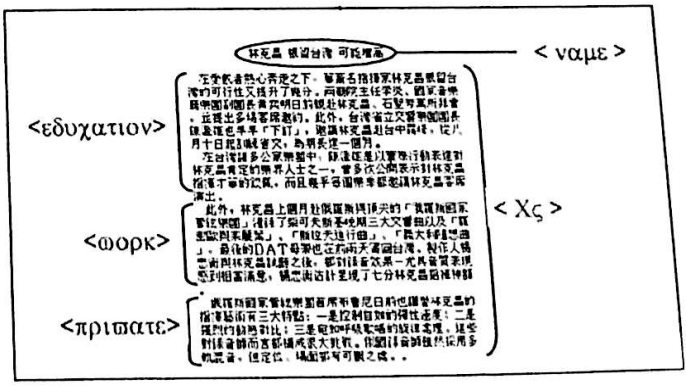
```
<module date="...">
      <title>...</title>
      <lecturer>
            <name>...</name>
      <weblink>...</weblink>
      </lecturer>
      <students>...</students>
</module>
```



- *DTD: describe the grammar and structure of permissible XML trees*

19

But What about this?



[Davies, 0

20

Print (1—100) numbers
without loop/recursion/goto

XML

- Meaning of XML-Documents is *intuitively* clear
  - due to "*semantic*" Mark-Up
  - tags are domain-terms
- But, computers do not have intuition
  - tag-names do not provide semantics for machines.
- DTDs or XML Schema specify the *structure* of documents, *not* the meaning of the document contents
- XML lacks a semantic model
  - has only a "surface model", i.e. tree

21

XML: limitations for semantic markup

- XML representation makes no commitment on:
  - Domain specific ontological vocabulary
    - Which words shall we use to describe a given set of concepts?
  - Ontological modelling primitives
    - How can we combine these concepts, e.g. "car is a-kind-of (subclass-of) vehicle"

⇨ requires pre-arranged agreement on vocabulary and primitives

- Only feasible for closed collaboration
  - agents in a small & stable community
  - pages on a small & stable intranet ☹
    .. not for sharable Web-resources

[Davies, 03]

22

XML is a first step

- Semantic markup
  - HTML ⇨ layout
  - XML ⇨ content
- Metadata
  - within documents, not across documents
  - *prescriptive*, not *descriptive*
  - No commitment on vocabulary and modelling primitives
- RDF is the next step

[Davies, 03]

23

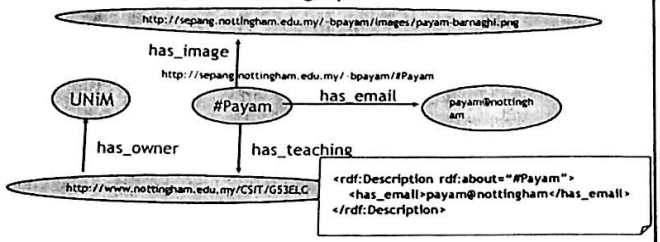
Resource Description Framework (RDF)

- A standard of W3C W3C RDF
- Relationships *between* documents
- Consisting of triples or sentences:
  - <subject, property, object>
  - <"Mozart", composed, "The Magic Flute" >
- RDFS extends RDF with standard "ontology vocabulary":
  - Class, Property
  - Type, subClassOf
  - domain, range

24

RDF for semantic annotation

- RDF provides metadata about Web resources
- Object -> Attribute-> Value triples
- It has an XML syntax
- Chained triples form a graph



RDF: Basic Ideas

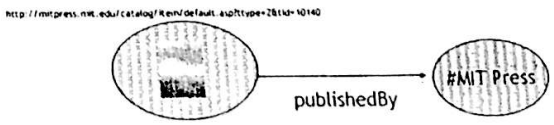
- Resources
  - Every resource has a URI (Universal Resource Identifier)
  - A URI can be a URL (a web address) or a some other kind of identifier;
  - An identifier does not necessarily enable access to a resources
  - We can think of a resources as an object that we want to describe it.
    - Books
    - Person
    - Places, etc.

RDF: Basic Ideas

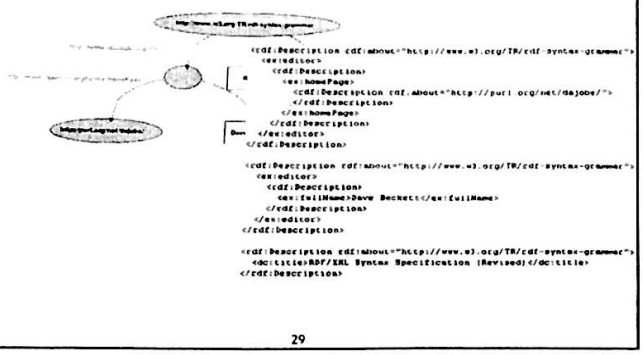
- Properties
  - Properties are special kind of resources;
  - Properties describe relations between resources.
  - For example: "written by", "composed by", "title", "topic", etc.
  - Properties in RDF are also identified by URIs.
  - This provides a global, unique naming scheme.

RDF: Basic Ideas

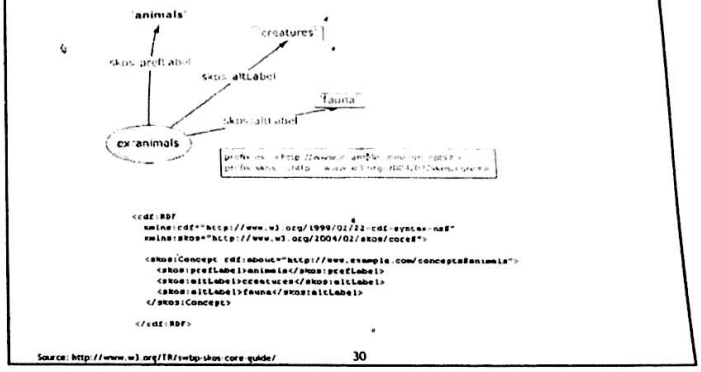
- Statements
  - A statement is an object-attribute-value triple.
  - It consists of a resources, a property, and a value.



RDF: Example



RDF Example



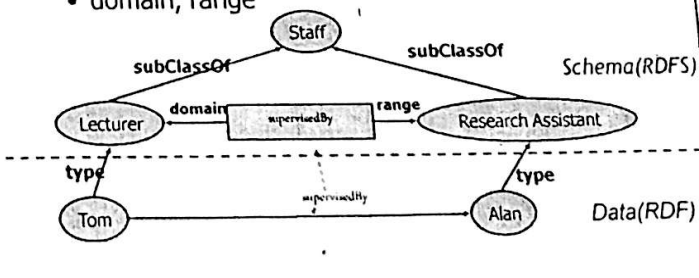
Source: http://www.w3.org/TR/swbp-skos-core-guide/

RDF Schema: Basic Ideas

- *RDF is a universal language that enables users to describe their own vocabularies.*
- *But, RDF does not make assumption about any particular domain.*
- *It is up to user to define this in RDF schema.*

What does RDF Schema add?

- Defines vocabulary for RDF
- Organizes this vocabulary in a typed hierarchy
  - Class, subClassOf, type
  - Property, subPropertyOf
  - domain, range



[adapted from: Studer et al, 04]

semantic web's SQL

"Sparkle"

Resource query language!

Querying RDF data

- Query Languages such as SPARQL, RQL.
- RDF is a directed, labeled graph data format for representing information in the Web.
- Most forms of the query languages contain a set of triple patterns.
- Triple patterns are like RDF triples except that each of the subject, predicate and object may be a variable.

33

Basic Queries

- The example provided in RQL.
- Using select-from-where
  - select specifies the number and order of retrieved data.
  - from is used to navigate through the data model.
  - where imposes constraints on possible solutions

34

Basic Queries: Example

select X,Y
From {X} writtenBy {Y}

X, Y are variables, {X} writtenBy {Y} represents a resource-property-value triple

35

Conclusions about RDF(S)

- Next step up from plain XML:
  - (small) **ontological commitment** to modeling primitives
  - possible to define **vocabulary**
- However:
  - **no** precisely described meaning
  - **no** inference model

[Davies, 03]

36

semantic web's SQL

"Sparkle"

Resource query language

Querying RDF data

- Query Languages such as SPARQL, RQL.
- RDF is a directed, labeled graph data format for representing information in the Web.
- Most forms of the query languages contain a set of triple patterns.
- Triple patterns are like RDF triples except that each of the subject, predicate and object may be a variable.

Basic Queries

- The example provided in RQL.
- Using select-from-where
  - select specifies the number and order of retrieved data.
  - from is used to navigate through the data model.
  - where imposes constraints on possible solutions

Basic Queries: Example

select X,Y
From {X} writtenBy {Y}

X, Y are variables, {X} writtenBy {Y} represents a resource-property-value triple

Conclusions about RDF(S)

- Next step up from plain XML:
  - (small) **ontological commitment** to modeling primitives
  - possible to define **vocabulary**
- However:
  - **no** precisely described meaning
  - **no** inference model

[Davies, 0

Ontologies

- The term ontology is originated from philosophy. In that context it is used as the name of a subfield of philosophy, namely, the study of the nature of existence.
- For the Semantic Web purpose:
  - "An ontology is an explicit and formal specification of a conceptualisation". *(R. Studer)*

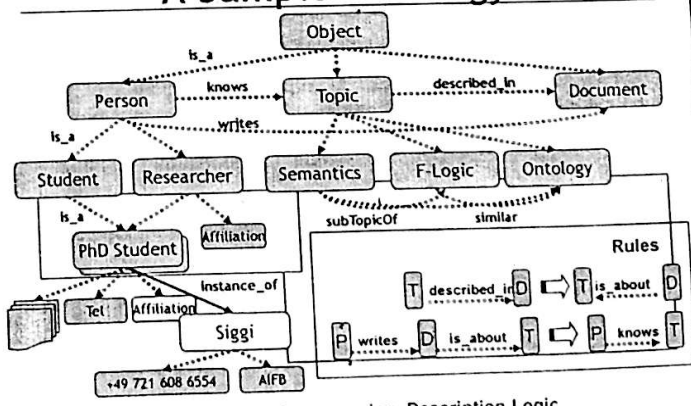
Ontologies and Semantic Web

- In general, an ontology describes formally a domain of discourse.
- An ontology consists of a finite list of terms and the relationships between the terms.
- The terms denote important concepts classes of objects) of the domain.
- For example, in a university setting, staff members, students, courses, modules, lecture theatres, and schools are some important concepts.

Ontologies and Semantic Web (cont'd)

- In the context of the Web, ontologies provide a shared understanding of a domain.
- Such a shared understanding is necessary to overcome the difference in terminology.
- Ontologies are useful for improving accuracy of Web searches.
- Web searches can exploit generalization/specialization information.

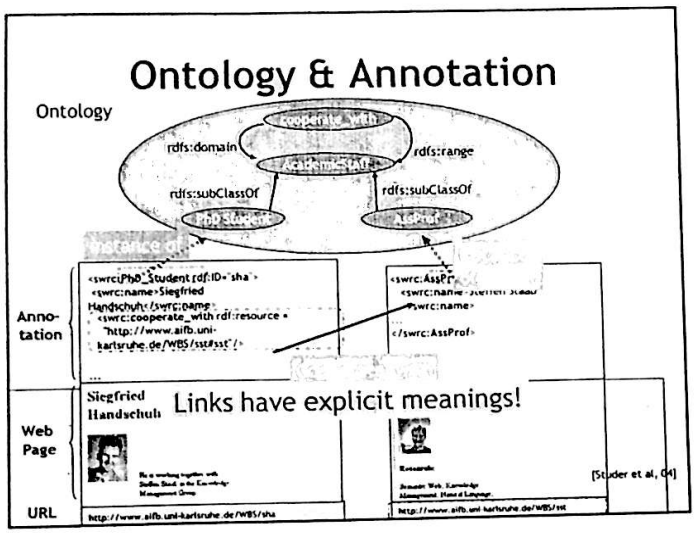
A Sample Ontology



- Major Paradigms: Logic Programming, Description Logic
- Standards: RDF(S); OWL

[Studer et al. 04]

Ontology & Annotation



Ontologies (OWL)

- RDFS is useful, but does not solve all possible requirements
- Complex applications may want more possibilities:
  - similarity and/or differences of terms (properties or classes)
  - construct classes, not just name them
  - can a program reason about some terms? E.g.:
    - "if «Person» resources «A» and «B» have the same «foaf:email» property, then «A» and «B» are identical"
  - etc.
- This lead to the development of OWL (Web Ontology Language)

source: Introduction to the Semantic Web, Ivan Herman, W3C

42

Ontology Languages for the Web

- RDF Schema is a vocabulary description language for describing properties and classes of RDF resources, with a semantics for generalization hierarchies of such properties and classes.
- OWL is a richer vocabulary description language for describing properties and classes.

43

OWL Language

- OWL is based on Description Logics knowledge representation formalism
- OWL (DL) benefits from many years of DL research:
  - Well defined semantics
  - Formal properties well understood (complexity, decidability)
  - Known reasoning algorithms
  - Implemented systems (highly optimised)
- Three species of OWL
  - OWL full is union of OWL syntax and RDF
  - OWL DL restricted to FOL fragment
  - OWL Lite is "easier to implement" subset of OWL DL
- OWL DL based on SHIQ Description Logic

[Davies, 03]

44

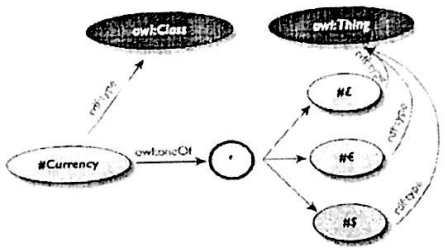
Classes in OWL

- In RDFS, you can subclass existing classes... that's all.
- In OWL, you can *construct* classes from existing ones:
  - enumerate its content
  - through intersection, union, complement
  - through property restrictions

source: Introduction to the Semantic Web, Ivan Herman, W3C

OWL classes can be "enumerated"

The OWL solution, where possible content is explicitly listed:



source: Introduction to the Semantic Web, Ivan Herman, W3C

Why develop an ontology?

- *To make define web resources more precisely and make them more amenable to machine processing*
- To make domain assumptions explicit
  - Easier to change domain assumptions
  - Easier to understand and update legacy data
- To separate domain knowledge from operational knowledge
  - Re-use domain and operational knowledge separately
- A community reference for applications
- To share a consistent understanding of what information means

[Davies, 03]

Ontology and Logic

- Reasoning over ontologies
- Inferencing capabilities

X is author of Y ⇨ Y is written by X

X is supplier to Y; Y is supplier to Z ⇨
X and Z are part of the same supply chain

Cars are a kind of vehicle;
Vehicles have 2 or more wheels ⇨
Cars have 2 or more wheels

[Davies, 03]

Logic and Inference

- Logic is the discipline that studies the principles of reasoning
- Formal languages for expressing knowledge
- Well-understood formal semantics
  - Declarative knowledge: we describe what holds without caring about how it can be deduced
- Automated reasoners can deduce (infer) conclusions from the given knowledge

source: A Semantic Web Primer, Grigoris Antoniou and Frank van Harmelen, MIT Press

An Inference Example

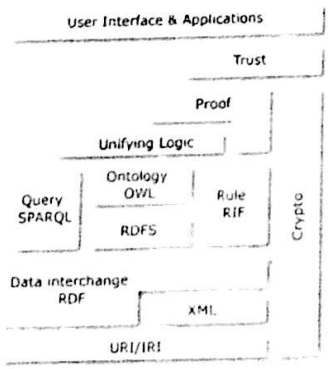
prof(X) → faculty(X)
faculty(X) → staff(X)
prof(michael)

We can deduce the following conclusions:

faculty(michael)
staff(michael)
prof(X) → staff(X)

source: A Semantic Web Primer, Grigoris Antoniou and Frank van Harmelen, MIT Press

Semantic Web Vision



Machine-processable, global Web standards:

- Assigning unambiguous names (URI)
- Expressing data, including metadata (RDF)
- Capturing ontologies (OWL)
- Query, rules, transformations, deployment, application spaces, logic, proofs, trust (in progress)

[Source: Emerging Web Technologies to Watch, Steve Bratt, W3C]

Semantic Web and AI?

- No human-level intelligence claims
- As with today's WWW
  - large, inconsistent, distributed
- Requirements
  - scalable, robust, decentralised
  - tolerant, mediated
- Semantic Web will make extensive use of current AI,
  - any advancement in AI will lead to a better Semantic Web
  - Current AI is already sufficient to go towards realizing the semantic web vision
- As with WWW, Semantic Web will (need to) adapt fast

[Davies, 03]

Semantic Web & Knowledge Management

- Organising knowledge in conceptual spaces according to its meaning.
- Enabling automated tools to check for inconsistencies and extracting new knowledge.
- Replacing query-based search with query answering.
- Defining who may view certain parts of information

Semantic Web Services

Web Services

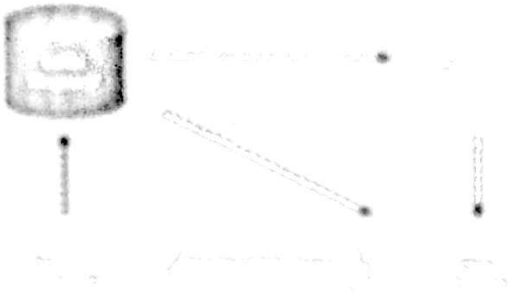
- Web Services provide data and services to other applications.
- Thee applications access Web Services via standard Web Formats (HTTP, HTML, XML, and SOAP), with no need to know how the Web Service itself is implemented.
- You can imagine a web service like a remote procedure call (RPC) which it returns a message in an XML format.

Web Services

- loosely coupled, reusable components
- encapsulate discrete functionality
- distributed
- programmatically accessible over standard internet protocols
- add new level of functionality on top of the current web

[Stollberg et al 05]

The Promise of Web Services



Deficiencies of WS Technology

- [illegible]
- [illegible]
  - [illegible]
  - [illegible]
  - [illegible]
  - [illegible]

[illegible]

Semantic Web Services

Semantic Web Technology

[illegible]

Web Service Technology

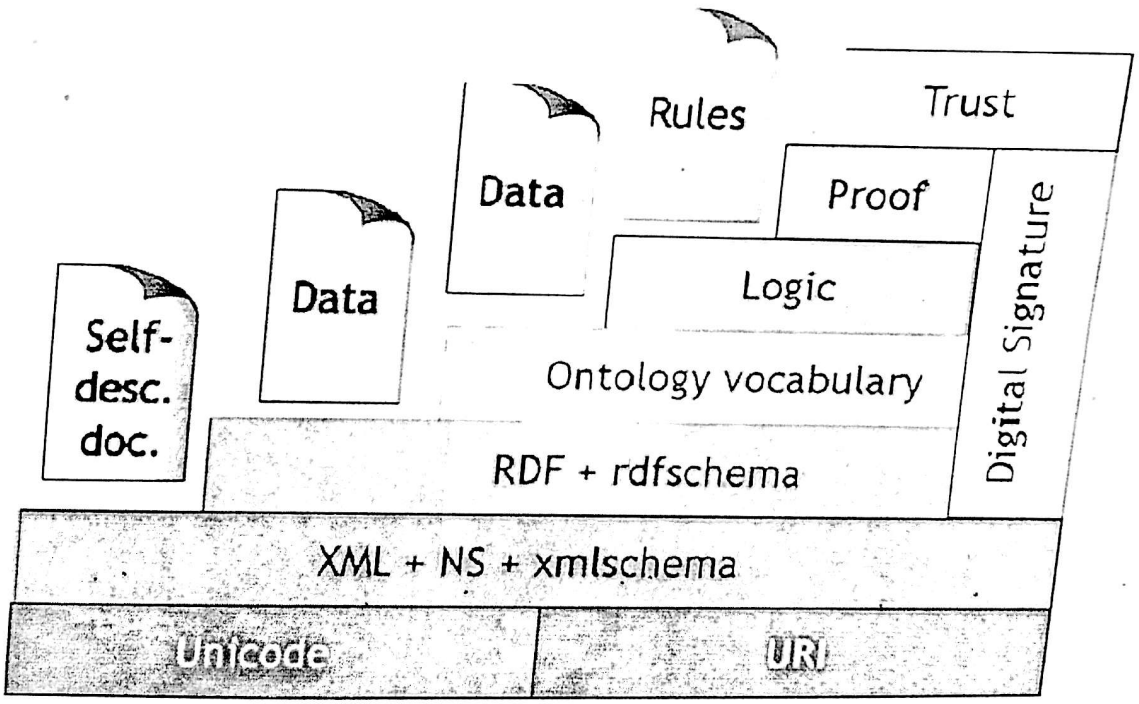
[illegible]

→ **Semantic Web Services as integrated solution for realizing the vision of the next generation of the Web**

Semantic Web Services

- define exhaustive description frameworks for describing Web Services and related aspects (**Web Service Description Ontologies**)
- support ontologies as underlying data model to allow machine supported data interpretation (**Semantic Web aspect**)
- define semantically driven technologies for automation of the Web Service usage process (**Web Service aspect**)

Semantic Web Layered Approach



Semantic Web Technologies

- Explicit ***metadata*** XML , RDF
- ***Ontologies*** to standardise concepts and relations between them
- ***Logic*** and ***Inference***: languages founded in various flavours of logic
- ***Software Agents***: make use of all the above to help us in our tasks

A more Comprehensive Representation

- XML-based representations are more easily processable by machines, since they are more structured

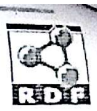
- XML based

```
<department>
    <departmentName>Artificial intelligence
    </departmentName>
    <hod>
      <name>Roger Right</name>
      <room>312</room>
      <telephone>23400007</telephone>
     <contactHr>11:30am-
    13:30pm</contactHr>
    </hod>
    <staff>
       <lecturer>Steve Runner</lecturer>
       <lecturer>George Cool</lecturer>
       <secretary>Mary Nice</secretary>
    </staff>
</department>
```

Ontologies

- The term ontology originates from philosophy:
  - The study of the nature of existence
- Ontology is the study of the categories of things that exist or may exist in some domain…it is a catalogue of the types of things that are assumed to exist in a domain ***D*** from the perspective of a person who uses a language ***L*** to talk about ***D***. *(Sowa 1997)*
- Think of an ontology as a vocabulary used to describe things *(Guarino 1998)*
- *Ontologies are used to facilitate knowledge sharing and reuse by formally defining a shared conceptualization*

Components of Ontologies



- An ontology describes formally a domain of discourse and includes the following components.
- **Terms** denote important concepts (or classes of objects) in the domain
  - e.g. professors, staff, students, courses, departments
- **Relationships** between these terms: most typical is a taxonomy relation (is-A)
  - a class C is a subclass of another class C' if every object in C is also included in C'
  - e.g. all professors are staff members

Other Ontology Components

- Properties:
  - e.g. X teaches Y
- Value restrictions
  - e.g. only faculty members can teach courses
- Disjointness statements
  - e.g. faculty members and general staff are disjoint
- Logical relationships between objects
  - e.g. every department must include at least 10 faculty members

Ontologies on the Web

- Ontologies are ideal to provide a shared understanding of a domain: enable semantic interoperability
  - overcome differences in terminology
  - issue: mappings between ontologies
- Ontologies are useful for the organization and navigation of Web sites
- Ontologies are useful for improving the accuracy of Web searches
  - search engines can look for pages that refer to a precise concept in an ontology

Semantic Web Languages

EXPRESSIVE

- Need languages to define ontologies
- Initially there where RDF/Schema:
  - Resource Description Framework
- then came DAML and OiL
- now we have a W3C recommendation for OWL
  - Web Ontology Language

Logic and Inference

- Logic is the discipline that studies the principles of reasoning
- Formal languages for expressing knowledge
- Well-understood formal semantics
  - Declarative knowledge: we describe what holds without caring about how it can be deduced
- Automated reasoners can deduce (infer) conclusions from the given knowledge

Machine understandable…

- Published facts
  - B related-to A
  - C related-to A
  - D related-to C
- Query
  - Return all entities related to A

    ?x related-to A
- Result
  - B
  - C

Machine understandable + inference

- Published facts
  - B related-to A
  - C related-to A
  - D related-to C
  - also declare that related-to is transitive
    ?x related-to ?y and ?y related-to ?z => ?x related-to ?z
- Query
  - Return all entities related to A
    ?x related-to A
- Result
  - B
  - C
  - D