

The World Wide Web...



- **Works reasonably well for single document texts, or for finding sites based on single document text**
 - ◆ Cannot integrate information from multiple documents
 - ◆ Cannot find things in databases, programs, devices and sensors
 - ◆ Cannot ever get better!
 - Keyword-based IR will never really do better than it does today (in satisfying user needs)



For Data we are still pre-web



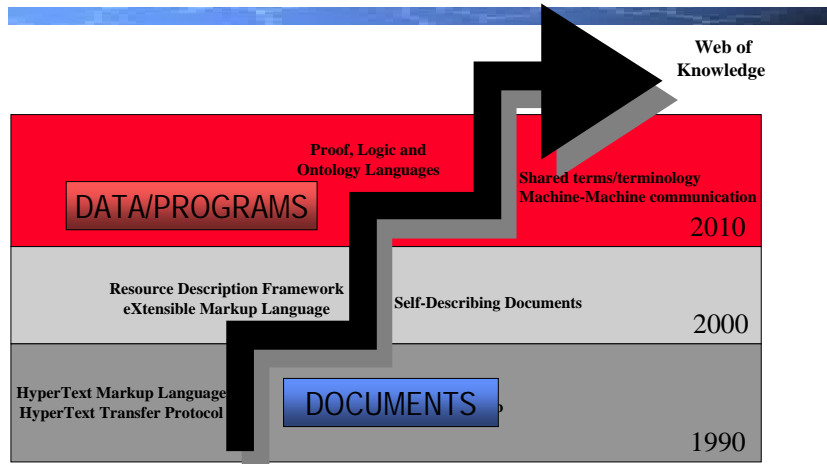
... Suppose you are browsing the web and you come across a web page about a meeting. It has the time and place and links to other documents including the home pages of other people involved in organizing and attending the meeting. You decide to attend, and click the "register" button. At this point, you would like your calendar to have an entry at the right date and time, with hypertext links to the details. You would like your in-car navigation system, at that date and time, to be programmed with the coordinates of the location. You would like your Rolodex to seem to contain, until the meeting is over, the contact info for the people involved. You'd like to do all this with one click.

.... **For data (and the programs that process it) we are still pre-Web!**

(Hendler, Berners-Lee, Miller 02)



The Evolving Web

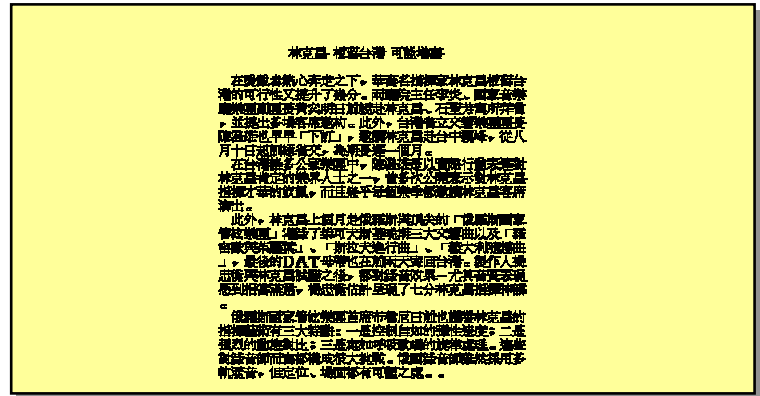


Based on Berners-Lee, Hendler; *Nature*, 2001

Brighton, Mar 2002

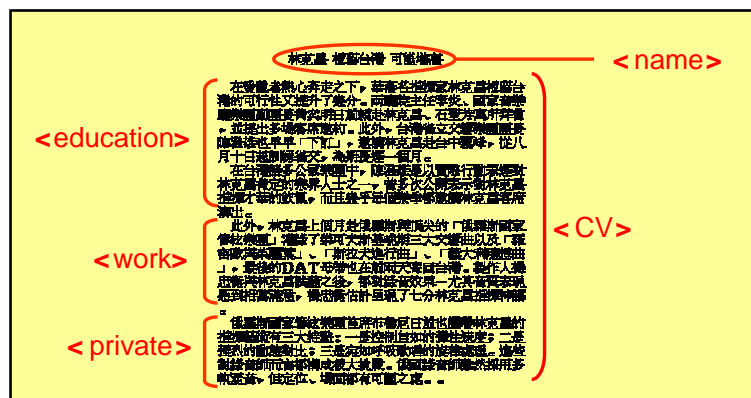
Why is this hard?

This is what a web-page in natural language looks like for a machine



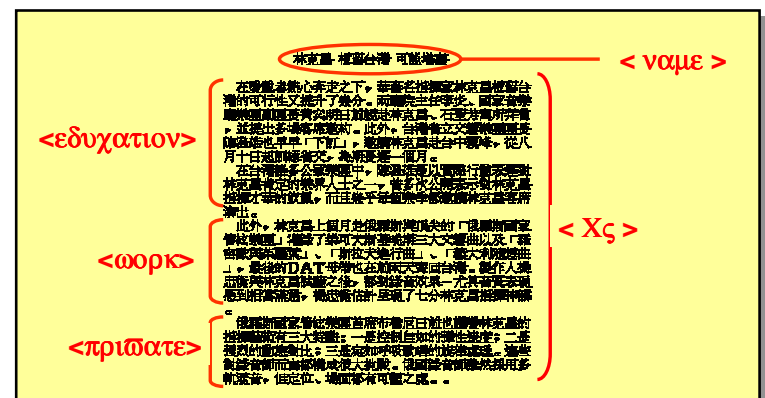
XML helps

XML allows "meaningful tags" to be added to parts of the text

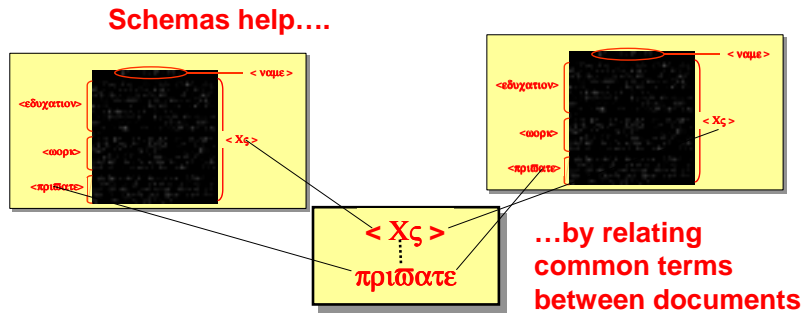


XML ≠ machine accessible meaning

But to your machine, the tags look like this....

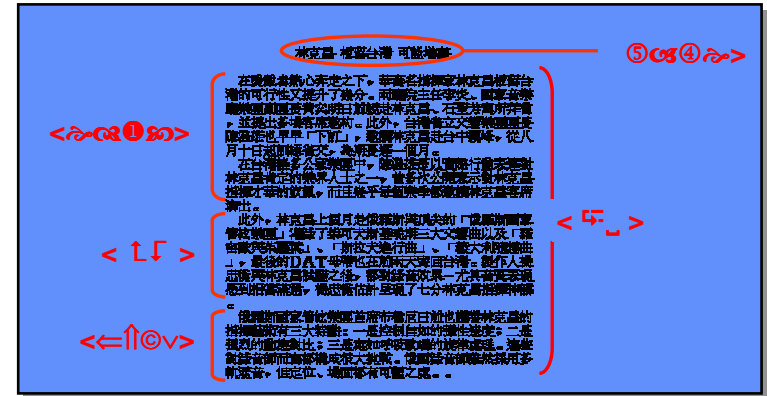


Schemas take a step in the right direction

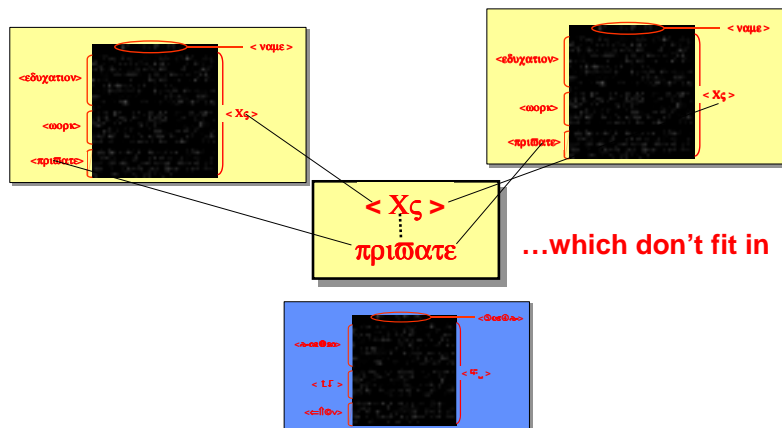


But other people use other schemas

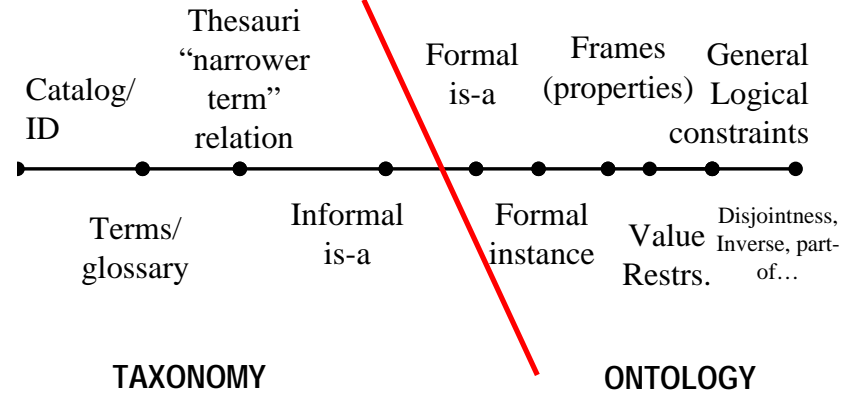
Someone else has one like this....



The "semantics" isn't there



AI can help



(McGuinness, 99)





AI on the web

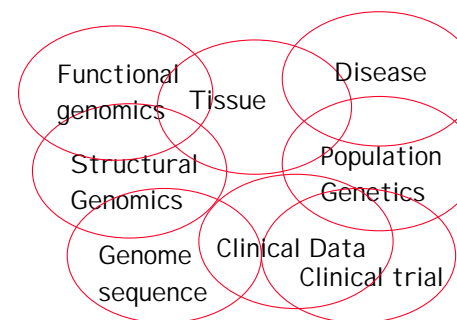
• Many characteristics of semantics, on the web, violate traditional AI/KR assumptions!

- ◆ It's Large and It Grows "organically"
- ◆ Lack of Referential Integrity
- ◆ High Variety in Quality of Knowledge
- ◆ Diversity of Content
- ◆ Unknown/unpredictable Use Scenarios for the Knowledge
- ◆ Problems of Trust, No Single Authority
- ◆ Knowledge acquired, not engineered

Brighton, Mar 2002



Common vocabulary?



(Genome World - from Goble, 01)

Brighton, Mar 2002



A distributed ontological representation

• Small communities define common semantics

- ◆ Technical Vocabularies abound
 - Mission specific
 - Technical jargons
 - Shared values

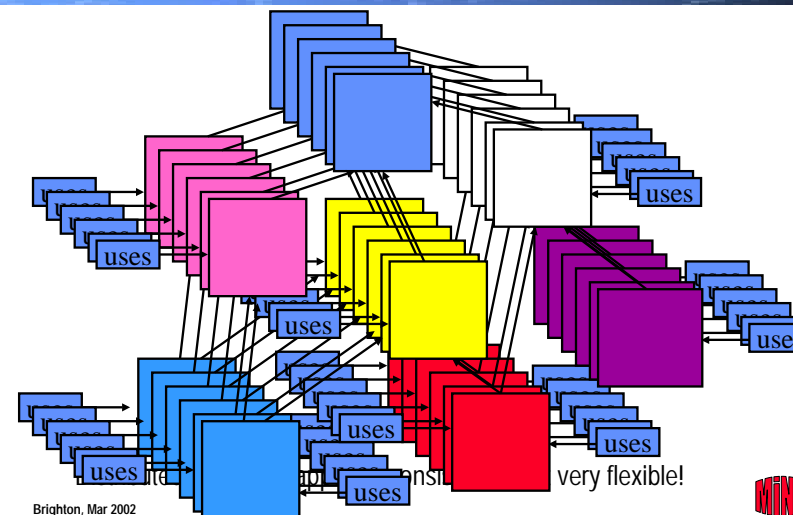
• Larger communities form around shared terms

- ◆ Mapping and "articulation" become crucial
 - Interoperability at web languages level
 - Top-Down (AIA defines critical aircraft properties) or bottom up (Oh, a "foxbat" is a Mig29)
 - Business case for improving communication!

Brighton, Mar 2002



This leads to a radically new view of interoperation



Brighton, Mar 2002



But, like the web...



Semantic Error 409- Ontology Not Found

You've encountered a "Ontology Not Found" error while trying to access a semantic term grounded on the [University of Maryland Computer Science Department](#) semantic web server.

Brighton, Mar 2002



Semantic Web "roadmap"



Brighton, Mar 2002



Semantic Web Today



The Semantic Web of 2002 resembles the early days of the World Wide Web

- ◆ Development funded primarily by Govt, but emerging corporate interest
- ◆ A lot of excitement, but confusion as to business case
- ◆ Open source tools and "geeks in control"
- ◆ Standards starting to stabilize to point where they permit deployment
- ◆ Developer tools, libraries, languages



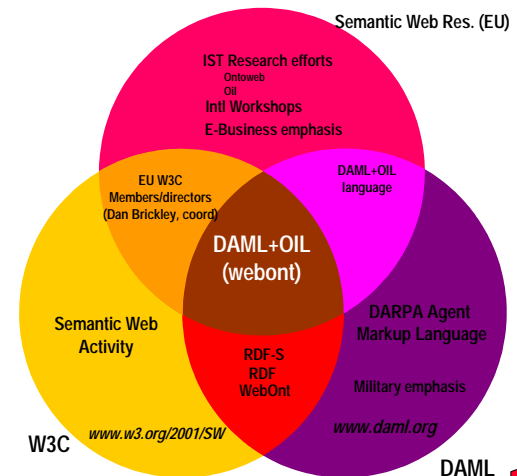
Brighton, Mar 2002



A number of organizations cooperating



- ◆ **US and EU Govt funding**
 - ◆ Both working closely with W3C to create a web standard
 - ◆ Works closely with EU on international acceptance
 - ◆ Joint tool development in open source arenas
- ◆ **Desire to include more partners**
 - ◆ Japan/Asia
 - ◆ Australia



2002 govt investment: US DoD = \$20,000,000; EU 20,000,000

Brighton, Mar 2002



W3C Web Ontology Working Group



- Web Ontology Working Group in the **W3C Semantic Web Activity** aimed at “extending the semantic reach of current XML and RDF meta-data efforts.”
- History
 - ◆ W3C Announcement in November 2001 - <http://lists.w3.org/Archives/Public/www-rdf-logic/2001Nov/0000.html>
 - ◆ Weekly teleconferences starting in November 2001
 - ◆ DAML+OIL is submitted as a [joint committee](#) effort published as a [W3C note](#) in December 2001.
 - ◆ First [Face to Face Meeting](#) in January 2001 - next meeting in Amsterdam in April, following meeting in US in July, UK in Oct.
 - ◆ First Working Draft (Use Case and requirements document) released 3/7/02.
 - <http://www.w3.org/TR/webont-req/>

Brighton, Mar 2002



Membership



- Current **Working** Group includes over 50 members from over 30 organizations.
 - ◆ Industry including:
 - Large companies such as Daimler Chrysler, EDS, Fujitsu, HP, Intel, Lucent, Nokia, Philips Electronics, Unisys, ...
 - Newer/smaller companies such as IVIS Group, Network Inference, Stilo Technology, Unicom Solutions, ...
 - ◆ Government and Not-For-Profits:
 - Defense Information Systems Agency, Interoperability Technology Association for Information Processing, Japan (INTAP) , Intelink Mgt Office, Mitre, ...
 - ◆ Universities and Research Centers:
 - University of Bristol, University of Maryland, University of Southampton, Stanford University, ...
 - DFKI (German Research Center for Artificial Intelligence), Forschungszentrum Informatik
 - ◆ Invited Experts
 - Well-known academics from non-W3C members
 - New experts joining from Medical Record and Digital Library Communities

Brighton, Mar 2002



Open Process (but still “Geek”y)



- The snowball is rolling
 - ◆ Joint development between DARPA/EU/and W3C communities
 - Archives of “Joint US/EU committee” available at <http://www.daml.org/>
 - ◆ Languages and tools are available to play with
 - <http://www.daml.org/>, <http://www.w3.org/2001/sw/WebOnt/>
 - ◆ W3C interest group available for those wishing to join the discussion
 - <mailto:Www-rdf-logic@w3c.org> (live or archived)
- Ongoing process
 - ◆ W3C Web Ontology Working Group (started Nov 1, 01)
 - All discussions/archives open to public
 - ◆ DARPA program (Murray Burke, Prog. Mgr)
 - All non FOUO materials on daml.org
 - ◆ W3C Interest Groups
 - www-rdf-interest@w3c.org, www-rdf-logic@w3c.org, www-rdf-rules@w3c.org
 - ◆ Ongoing DoD and commercial projects
 - Horus deployed in US Intelligence application
 - Large company labs, small startups emerging

Brighton, Mar 2002



www.daml.org



- Language Specifications
- DAML Newsletter (you can subscribe)
- Collection of web tools
 - ◆ Primarily for developers not end users
- Ontology library
 - ◆ 175+ ontologies
- DAML crawler
 - ◆ over 17,000 pages w/2,900,000+ DAML statements
- DAML page use
 - ◆ About 2,000,000 hits to date (under 2 years)

Brighton, Mar 2002



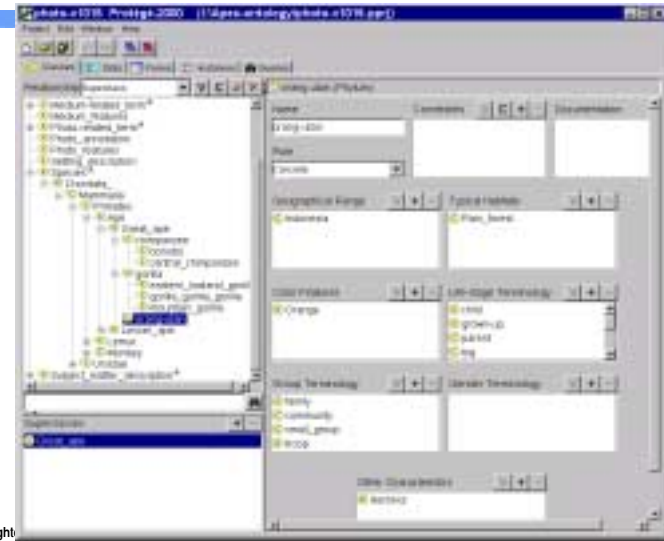
Semantic Web futures



Brighton, Mar 2002



Animal ontology



Brighton, Mar 2002



Making Markup Easier



Brighton, Mar 2002



Machine worries about the syntax



Brighton, Mar 2002



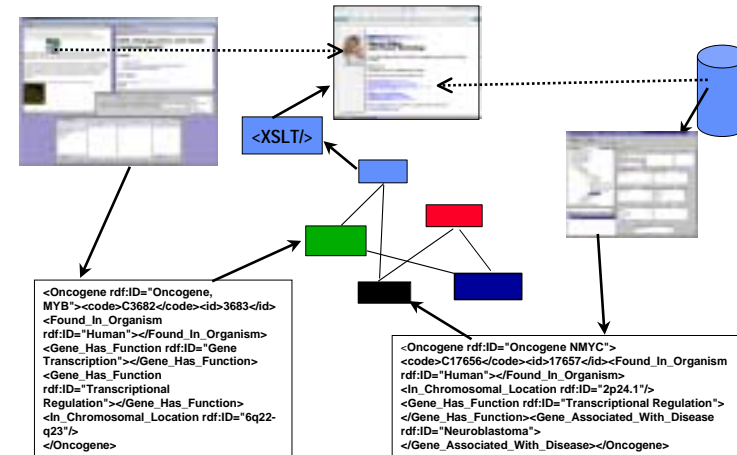
Use that markup in query/portal interfaces



Brighton, Mar 2002



Creating a "virtual" portal



Brighton, Mar 2002



Web "travel agents"



How many cows are there in Texas?

Query processed: 73 answers found

- ◆ [Google](#) document search finds 235,312 possible page hits.
- ◆ <http://www.../CowTexas.html> claims the answer is 289,921,836
- ◆ A database entitled "[Texas Cattle Association](#)" can be queried for the answer, but you will need "authorization as a state employee."
- ◆ A computer program that can compute that number is offered by the State of Texas Cattleman's Cooperative, [click here](#) to run program.
- ◆ ...
- ◆ The "sex network" can answer anything that troubles you, [click here](#) for relief...
- ◆ The "UFO network" claims the "all cows in Texas have been replaced by aliens"



Brighton, Mar 2002



Going Beyond Text!



Query processed:

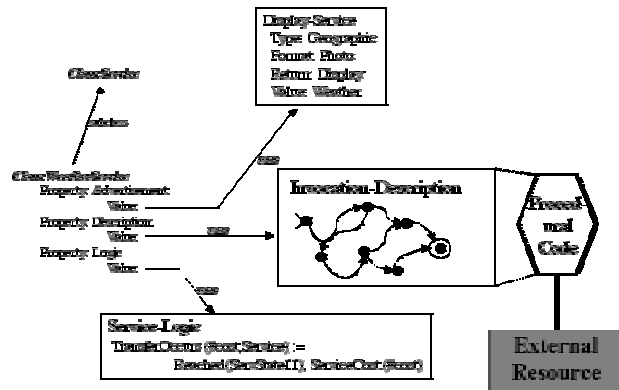
- ◆ A satellite image taken yesterday at 10 AM is available on the web at [http://...](#)
- ◆ A new satellite image, to be taken today at 10AM, will be available for \$100 — [click here](#) to authorize transfer of funds and obtain image (you will need a valid credit card number from one of the following providers: ...)
- ◆ In an emergency situation, a Coast Guard observer plane can be sent to any location within the area you indicate. Service Note: You will be responsible for cost of flight if the situation does not result in emergency pickup. [Click Here for more information.](#)
- ◆ A high altitude observer can be sent to your location in 13 hours. [Click here](#) to initiate procedure. (You will need to provide US military authorization, A valid military unit code, and the name of commanding officer)
- ◆ A service entitled "commercial service for providing satellite images" is advertised as becoming available in 2004. See [http://...](#) for more information



Brighton, Mar 2002



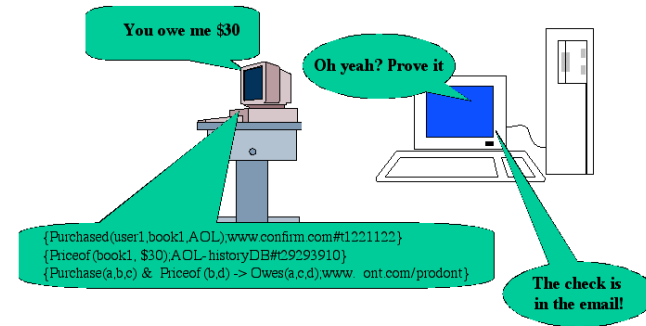
Service Descriptions



Brighton, Mar 2002



Web Logics



Brig



Web of trust



• Inference rules can be used to determine the credibility of claims

- ◆ We might believe any statement made by a reliable Newspaper
 - believe(x) :- claims(src, x) ^ reliableNewspaper(src)
- ◆ If we establish the Washington Post as reliable...
 - isa(http://www.washingtonpost.com, reliableNewspaper)
- ◆ or if we infer it
 - reliableNewspaper(x) :- linkto("http: ...", x)
 - reliableNewspaper(x) :- claims(src, y) ^ trusted(src) ^ predicate(y, reliableNewspaper) ^ arg(y, x)

Brighton, Mar 2002



Validating Web Pages



- Other claims might only be believed if there is supporting evidence from another source
 - ◆ We only believe that someone is a professor at a university if the university also claims that person is a professor, and the university is accredited

believe(c1) :- claims(x, c1) ^ predicate(c1, professorAt) ^ arg1(c1, x) ^ arg2(c1, y) ^ claims(c2, y) ^ predicate(c2, professorAt) ^ arg1(c2, x) ^ arg2(c2, y) ^ AccreditedUniversity(y)
AccreditedUniversity(u) :- link-from("http://www.cs.umd.edu/university-list", u)

Notice this one

Brighton, Mar 2002





● Buy into your favorite rule set

- ◆ believable(x) :- claims(src,x) ^
 accreditedbyChristianCoalition(src)
- ◆ believable(x) :- claims(src,x) ^
 linkfromMomsPage(src)
- ◆ believable(x) :- claims(src,x) ^
 accreditedby("http://foo.com/Unabomber/Friends/rules",SRC) ^
 Not-accreditedbyChristianColation(x)

Brighton, Mar 2002



● What about *human intelligence*?

- ◆ It's Large and It Grows Organically
- ◆ Lack of Referential Integrity
- ◆ High Variety in Quality of Knowledge
- ◆ Diversity of Content
- ◆ Unknown/unpredictable Use Scenarios for the Knowledge
- ◆ Problems of Trust, No Single Authority
- ◆ Knowledge acquired, not engineered

● Many characteristics of *human intelligence* violate traditional AI assumptions!

Brighton, Mar 2002



- It is no longer a question of whether the semantic web could come into being, it can and will
- We're already well past the starting gate
 - ◆ Web ontologies, term languages, "shims" to DB and services, research in proofs/rules/trust
- The "business case" is starting to emerge
 - ◆ Cross schema document mapping <http://www.daml.org>
 - ◆ Databases integrated into the web <http://www.semanticweb.org>
 - ◆ Service discovery and composition <http://www.w3.org/2001/sw>
 - ◆ Trust, Authentication, Accreditation
- The current environment is open, encouraging, moving fast, and exciting as heck
 - ◆ Come play!

Brighton, Mar 2002

