Towards a Science of the Web

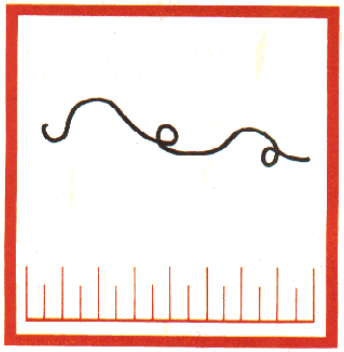
Jim Hendler

Tetherless World Senior Constellation Chair

Rensselaer Polytechnic Institute

Rensselaer

Today's Outline



Follow a thread,

into a ball of yarn,

and examine one knot…

…raising more questions than I answer.

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 cooperation. (Berners-Lee et al, 01)

If HTML and the Web made all the online documents look like one huge book, RDF, schema, and inference languages will make all the data in the world look like one huge database (Berners-Lee, 99)

The Semantic Web will enable machines to comprehend semantic documents and data, not human speech and writings. (Berners-Lee et al, 01)

Some standard quotes

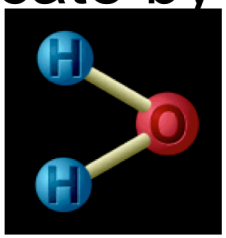
Tomorrow's talk

- The Dark Side of the Semantic Web
  - Talk and Q/A full of heresies against the "standard vision"

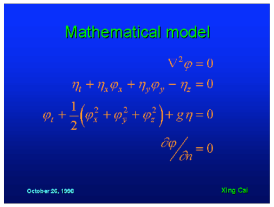
  - 11AM: CISA/AIAI, Room 4.03. Appleton Tower

A very old idea in new clothes

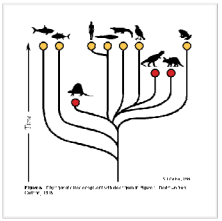
- Scientists communicate by use of models
  - c.f. Physical



  - c.f. Mathematical



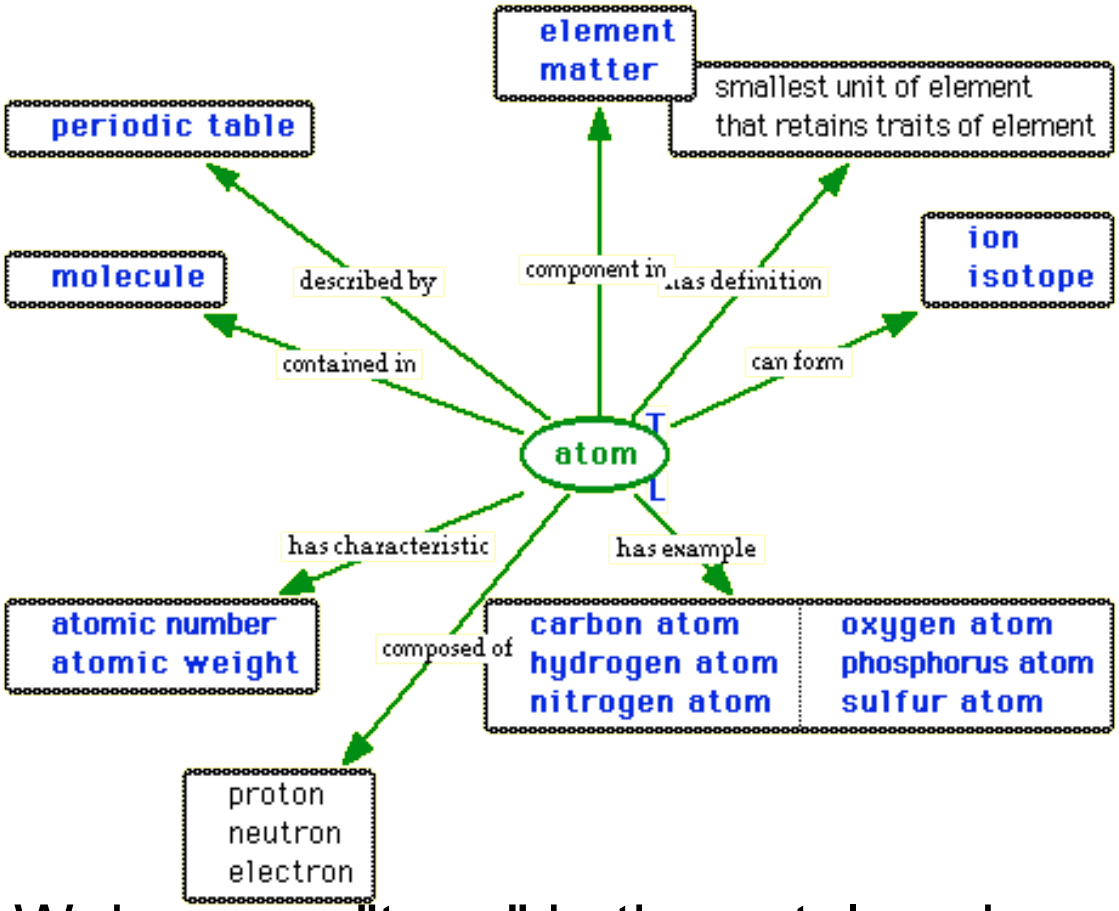
  - c.f. Organizational



Models expose semantics

Rensselaer

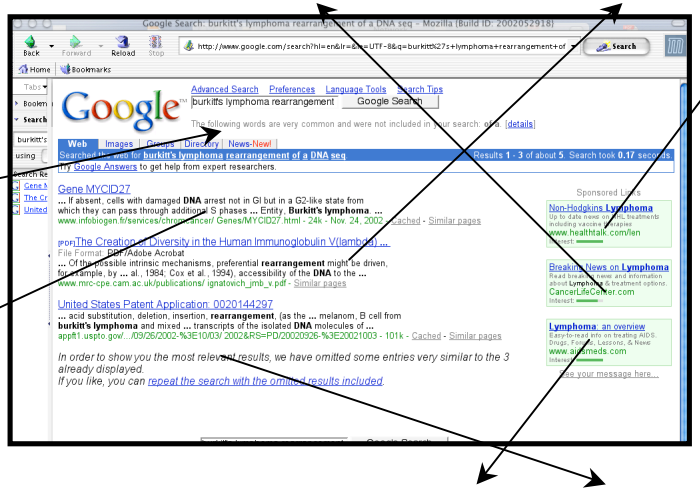
Ontology



Sem Web: every "term" in the ontology becomes a uniquely named, Web linkable, concept

Adding ontology

Burkitt's Lymphoma



PVT

Rearrangement of a DNA sequence homologous to a <cell-type>cell-virus junction fragment </cell-type>in several <disease>Moloney murine leukemia</disease> virus-induced <organism>rat</organism> thymomas

PubMed

Semantic Web

Oncogene(MYC):
Found_In_Organism(Human).
Gene_Has_Function(Transcriptional_Regulation).
Gene_Has_Function(Gene_Transcription).
In_Chromosomal_Location(8q24).
Gene_Associated_With_Disease(Burkitts_Lymphoma).

Burkitt's Lymphoma

8q24
PVT1

PVT
Rearrangement of a DNA sequence homologous to a cell-virus junction fragment in several Moloney murine leukemia virus-induced rat thymomas

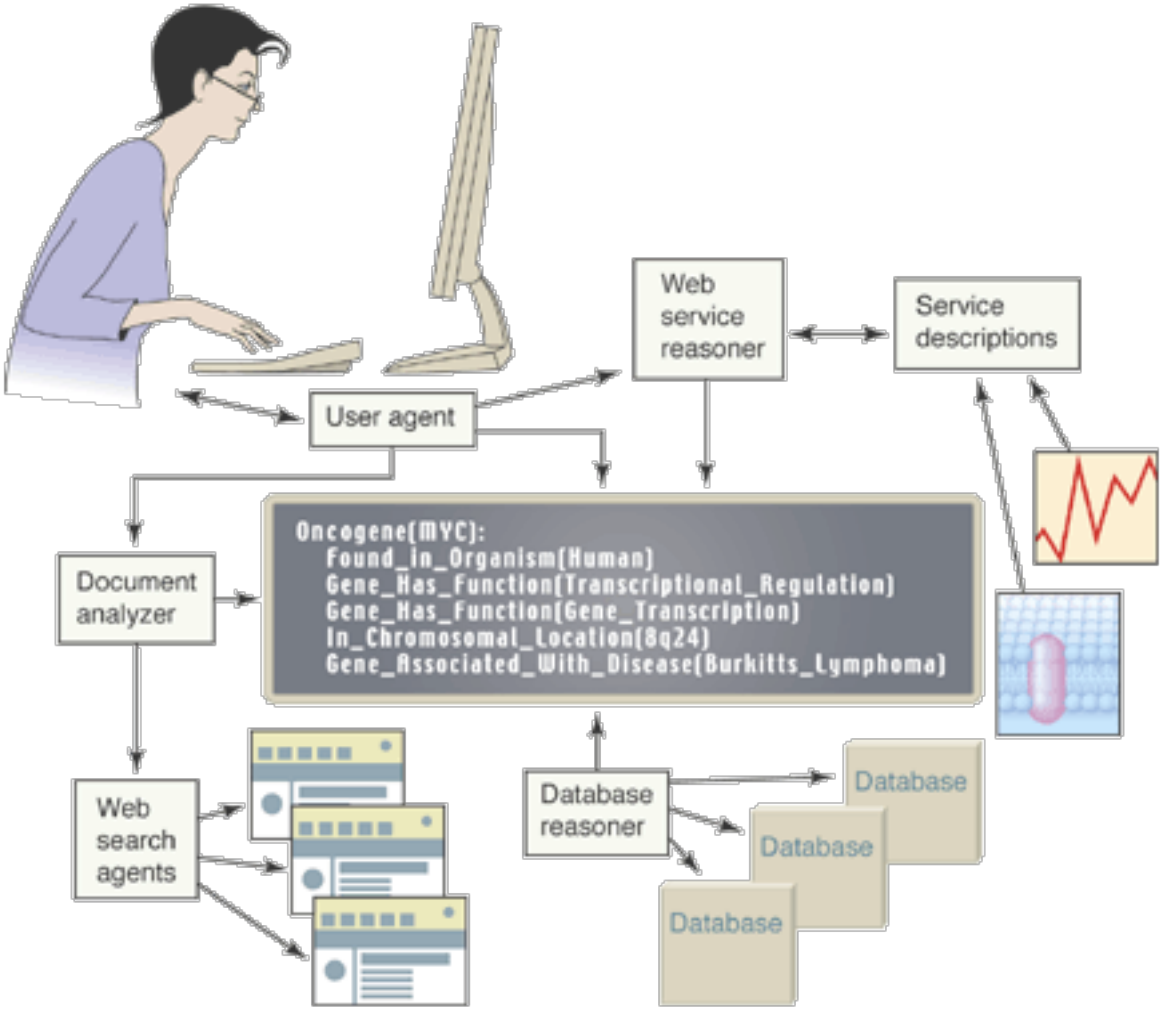
PubMed

Web Modeling Languages

- **Resource Description Framework (RDF)**
  - Few, but important, constraint
  - A basic, extensible assertional language
- **RDF Schema (RDFS)**
  - Weak structuring of sets of terms (taxonomy-esque)
  - Class and property hierarchies
  - Domain and Range constraints
- **The Web Ontology Language, OWL**
  - Stronger structuring of sets of terms (ontologies)
  - Everything in RDFS plus
    - Complex Class constructors (unionOf, intersectionOf)
    - Additional property features (inverse, transitive)
    - Class local property type and cardinality constraints
    - And more

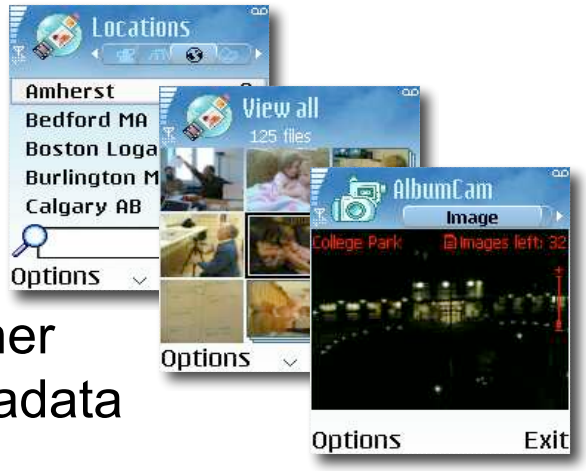
Using the links

- These models allow linking of
  - multimedia
  - databases
  - services
    - web
    - Grid?
  - meta-data repos
- Or any other Web resource!



Rensselaer

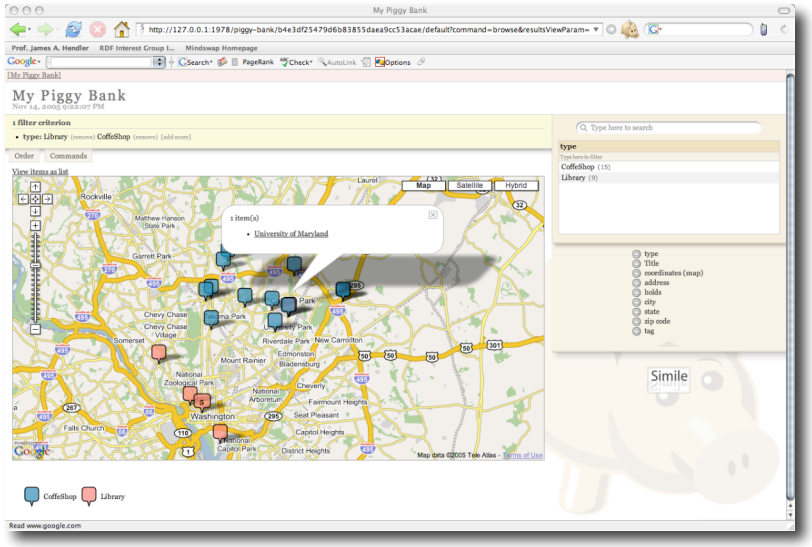
"Corporate Semantic Web", Gartner "hot pick" for 2006



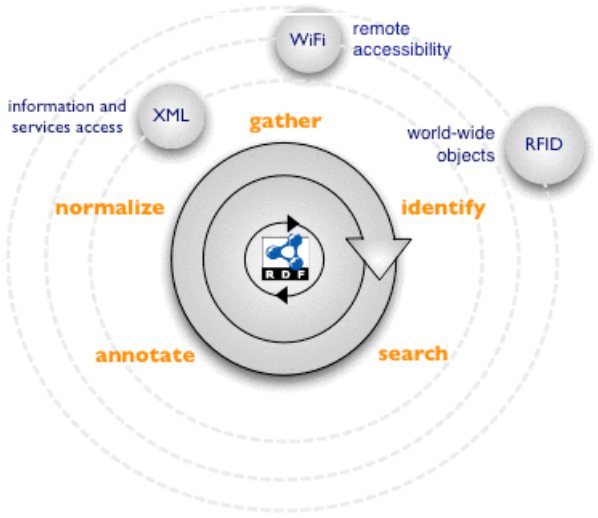
Richer metadata

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
      <rdf:Description rdf:about=""
          xmlns:pdf="http://ns.adobe.com/pdf/1.3/">
         <pdf:Producer>Acrobat Distiller 7.0.5 for Macintosh</pdf:Producer>
      </rdf:Description>
      <rdf:Description rdf:about=""
            xmlns:dc="http://purl.org/dc/elements/1.1/">
         <dc:format>application/pdf</dc:format>
         <dc:creator>
            <rdf:Seq>
               <rdf:li>James Hendler</rdf:li>
            </rdf:Seq>
         </dc:creator>
         <dc:title>
            <rdf:Alt>
               <rdf:li xml:lang="x-default">XMLideas.ppt</rdf:li>
            </rdf:Alt>
         </dc:title>
      </rdf:Description>
      <rdf:Description rdf:about=""
            xmlns:xapMM="http://ns.adobe.com/xap/1.0/mm/">
         <xapMM:DocumentID>uuid:93277c40-5534-11da-a3f2-000a95d6b344</xapMM:DocumentID>
         <xapMM:InstanceID>uuid:9327882b-5534-11da-a3f2-000a95d6b344</xapMM:InstanceID>
      </rdf:Description>
   </rdf:RDF>
```

Embedded meta-data



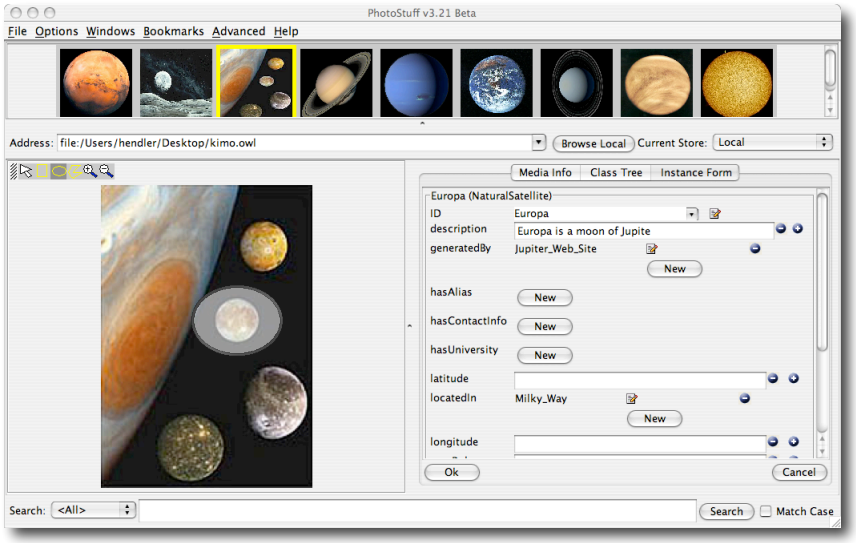
Data harvesting & visualization



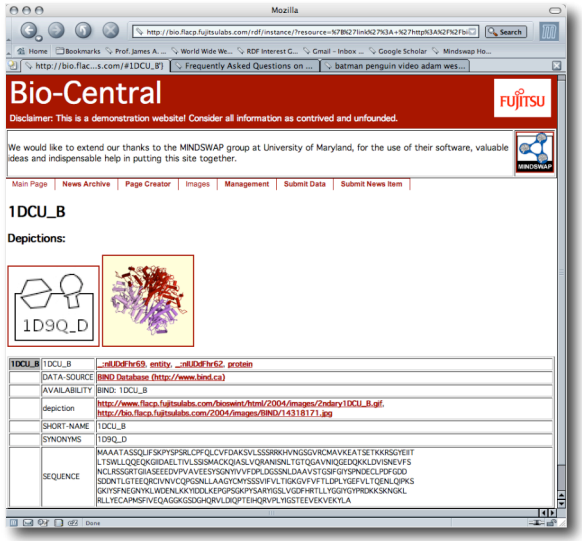
A Semantic Enterprise Information Integration Life Cycle

Enterprise data integration

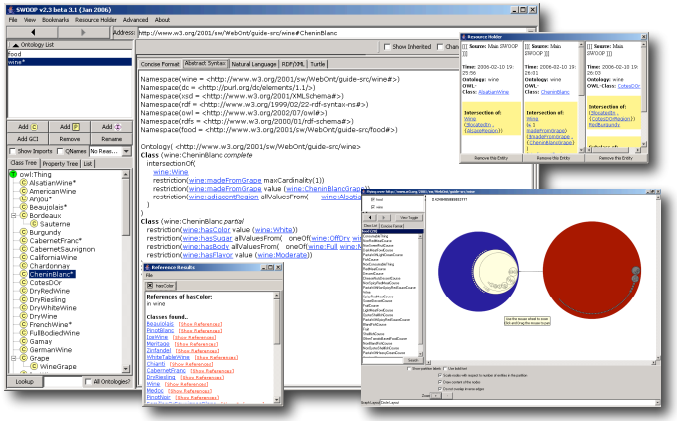
Rensselaer



Digital asset management



Semantic Web portals



Ontology editors (and other tools)

Semantic Web and social networking

Rensselaer

Transitioning Technologies

- Semantic Web Services
  - crucial for linking “programs” into the mix
  - “plumbing” for agents…
- Scaling Semantic Web stores to database sizes
- Information extraction and semantics ("Web 3.0")
  - can we “retrofit” semantics on the existing Web?
- Semantic Web information creation
  - can we make it so we don't have to retrofit in the future?
    - tools that help embed the semantics as a document is created
    - better dynamic integration of structured data into the Semantic Web

Significant Corporate Activity

- Semantic (Web) technology companies starting & growing
  - Siderean, SandPiper, SiberLogic, Ontology Works, Intellidimension, Intellisophic, TopQuadrant, Data Grid, …
- Bigger players buying in
  - Adobe, Cisco, HP, IBM, Nokia, Oracle, Sun, Vodaphone… announcements/use in 2005-2006
  - integrator and contractor uptake: Northrop Grumman buys Tucana, Lockheed-Martin uses SiberLogic in FCS, WebMethods buys Cerebra, …
  - tools being announced: AllegroGraph, TopBraid, …
- Government projects in and across agencies
  - US, EU, Japan, Korea, China, …
- Life sciences/pharma an increasingly important market
  - Health Care and Life Sciences Interest Group at W3C
- Many open source tools available
  - Kowari, RDFLib, Jena, Sesame, Protégé, SWOOP, Onto(xxx), Wilbur, …

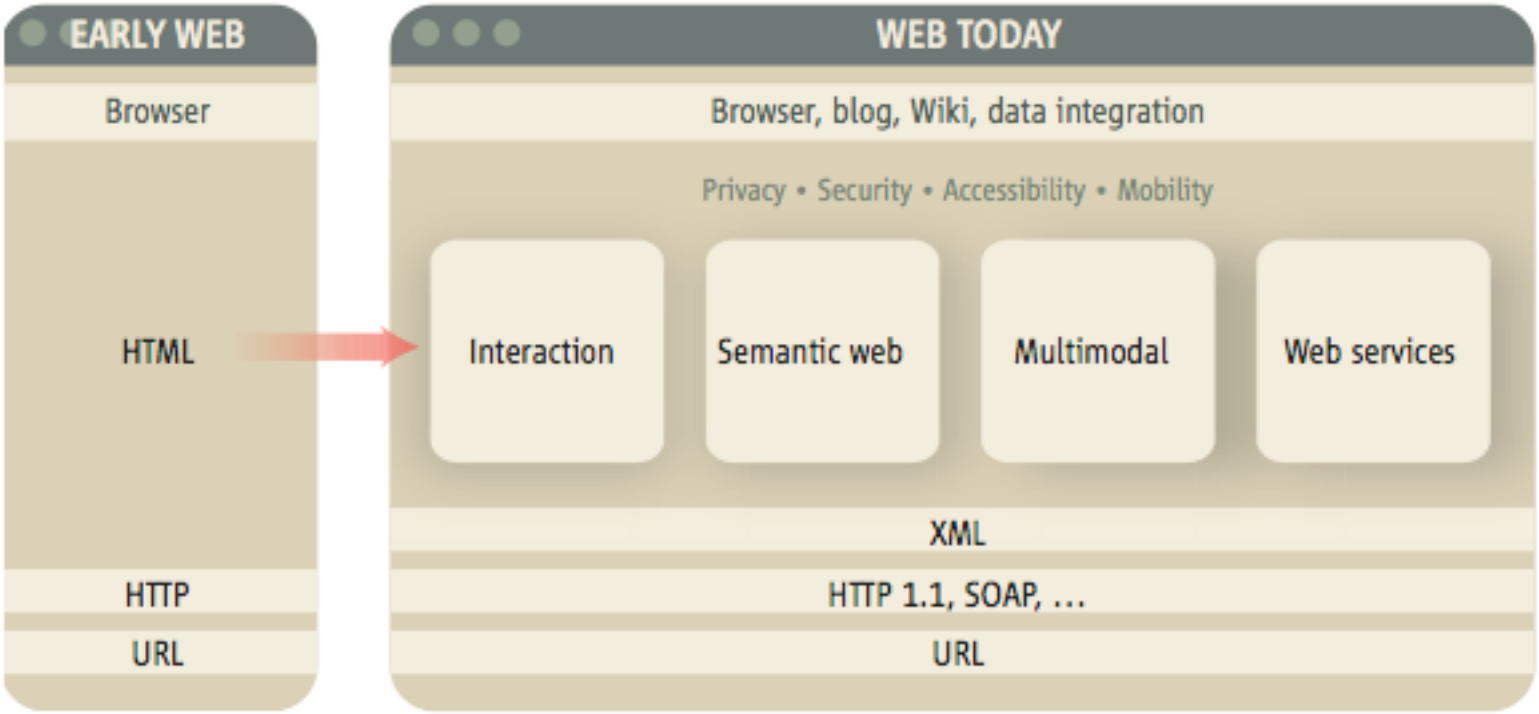
Universities should not be "competing" for low hanging fruit – so what comes next?

Towards a Science of the Web

- The Web has been transformational
  - Largest human information construct
- Understanding it requires a "science" that takes the Web as a primary area of interest
  - Math/Science approach to understanding the Web
  - Engineering approaches for its future
    - Semantic Web is an example
  - Understanding how to ensure its social benefit
- This requires a new interdisciplinary effort
  - Computer Science
  - Cognitive Science
  - Social/Policy issues

A new challenge for the Computing Field

Web Evolution



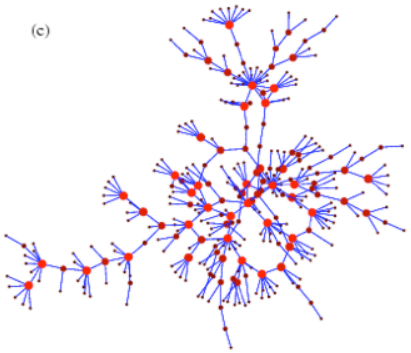
The HTTP/HTML (Rest) Web has evolved in complex ways

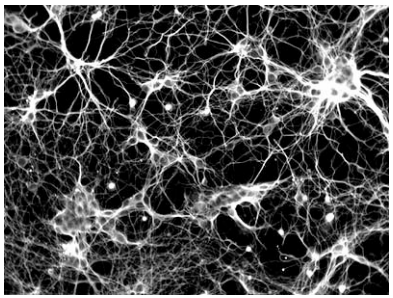
What we need to understand

- The features that make the web work
- The threats of fragmentation
- The macro from the micro
  - Think Google
- How to incorporate semantics
- How to exploit statistics
- How to build services
- How to instil trust, resilience and dependability

A "Systems" Science

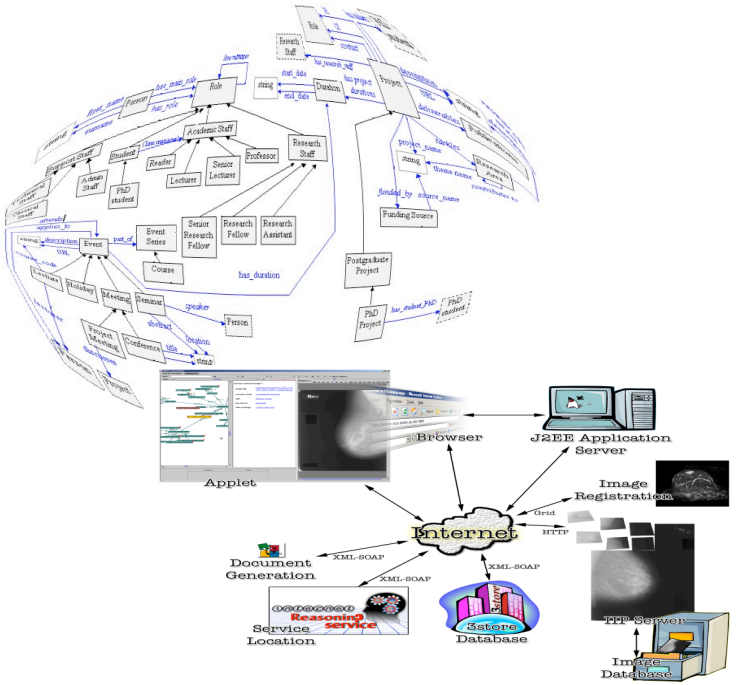
- The Web is so complex it needs its own Systems Science
- Inspiration from new formal models
- Inspiration from biology?





New Engineering

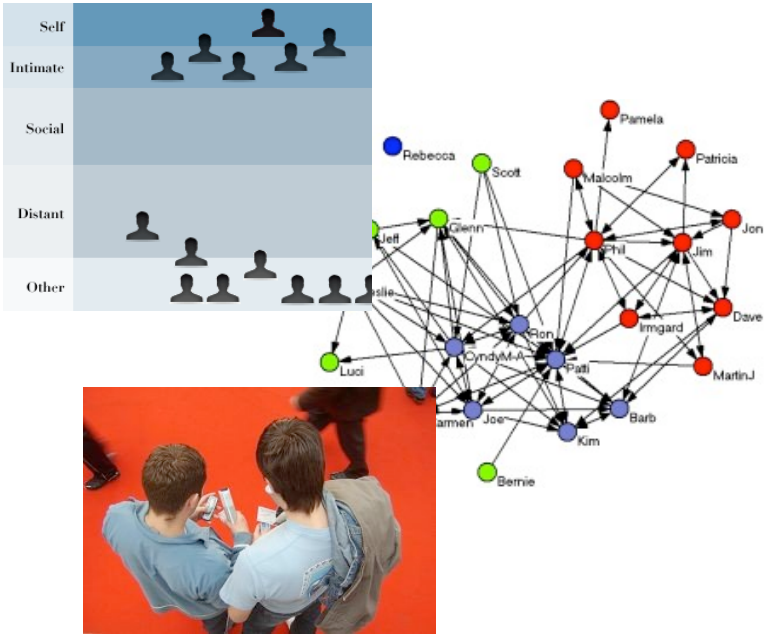
- The Web needs new engineering standards, methods and tools
- Access to data globally
- Decentralization for resilience



Rensselaer

New Social Models

- The Web is essentially social
  - And getting moreso
    - Facebook,myspace
    - Youtube, flickr
- The Web has huge impact on society
  - Information access and use
  - Privacy issues
  - Information flow control



Let's explore this thread a bit…

Rensselaer

Social impacts of the Web

- The "democratization" of trust, privacy, and other "social" issues provide many societal challenges
  - Some we're already seeing
    - cf. Blogs vs. the established media
    - cf. Information control by govt *vs.* personal freedoms
  - Many we're just starting to encounter
    - How to establish credibility in an open and distributed world
    - How to control private information in an increasingly transparent world
  - Some looming long term
    - cf. issues of virtual presence and online lives

Example:Web-Based Social Networks

- Websites and interfaces that let people maintain browsable lists of friends
- Last count
  - 142 social networking websites
  - Over 200,000,000 accounts
  - Full list at http://trust.mindswap.org
- Over 15,000,000 accounts are represented in FOAF, an OWL ontology

Adding "Trust"

- A number of these networks allow people to rate how much they trust others
  - Golbeck (2005) inferring trust from given ratings

$$t_{is} = \frac{\sum_{j \in adj(j) \mid t_{ij} \geq max} t_{ij} t_{js}}{\sum_{j \in adj(j) \mid t_{ij} \geq max} t_{ij}}$$

- With direct knowledge or a recommendation about how much to trust people, this value can be used as a filter in many applications
- Since social networks are so prominent on the web, it is a public, accessible data source for determining the quality of annotations and information

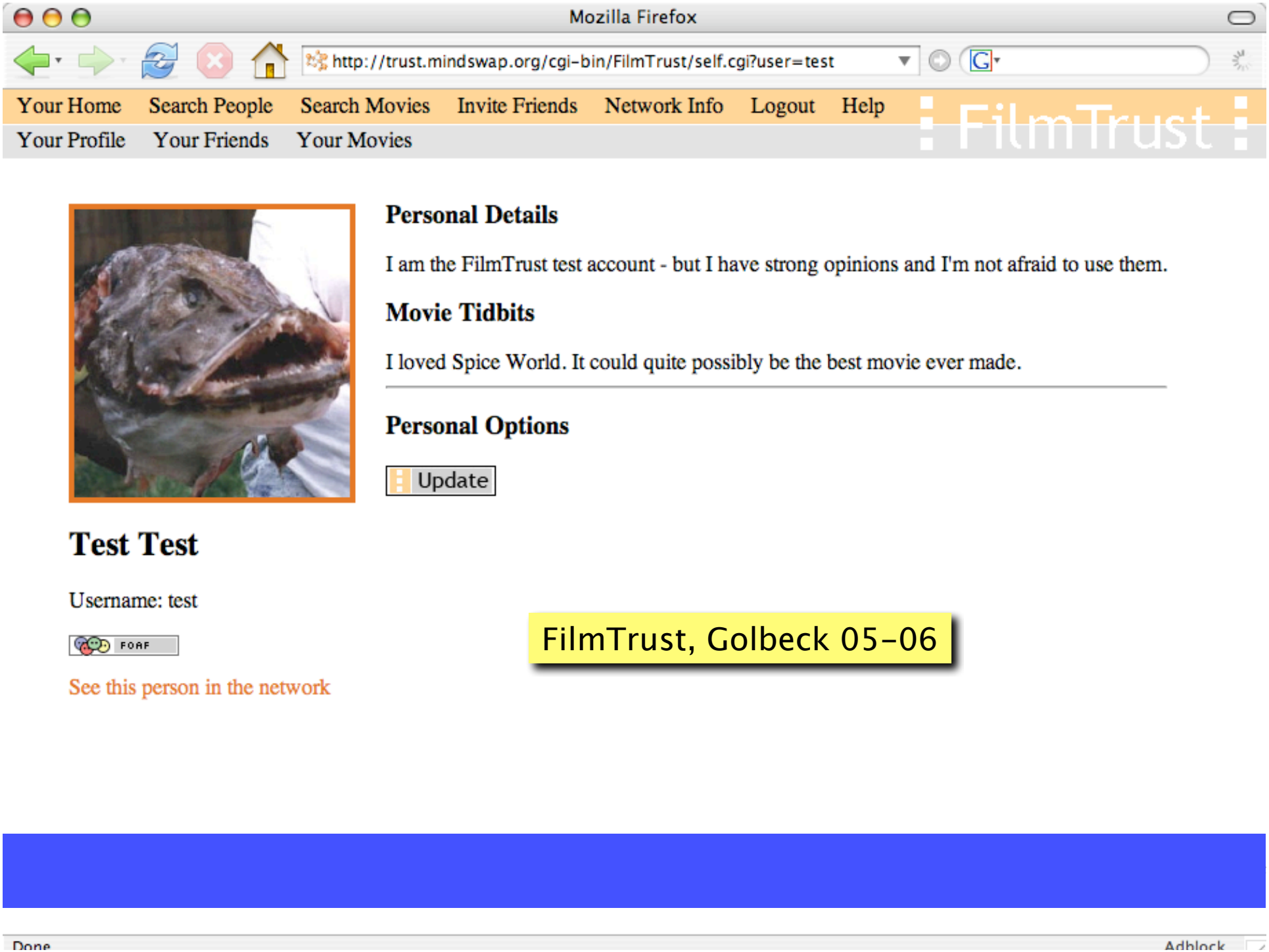
Trust Algorithm

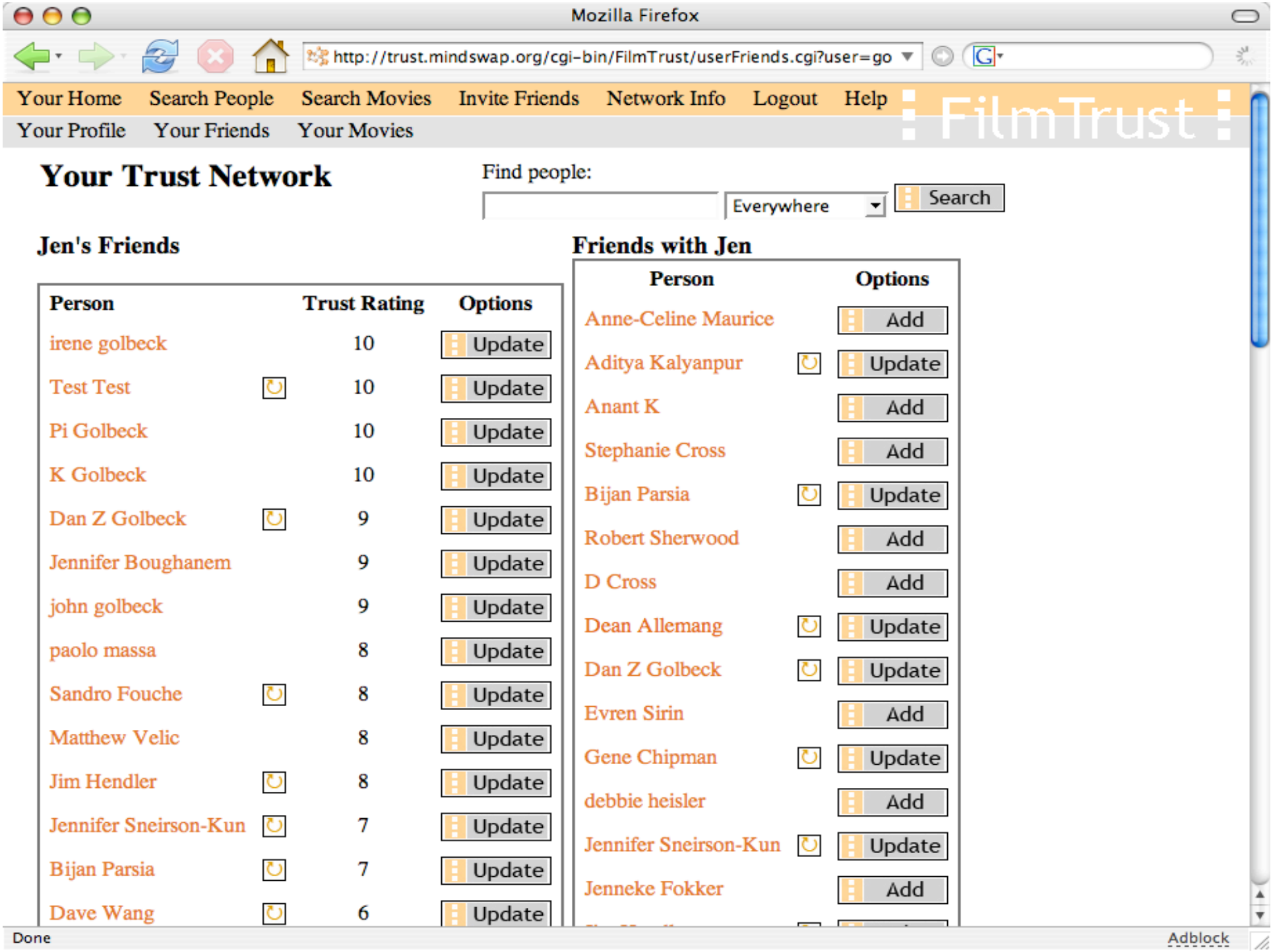
- If the source does not know the sink, the source asks all of its friends how much to trust the sink, and computes a trust value by a weighted average
- Neighbors repeat the process if they do not have a direct rating for the sink

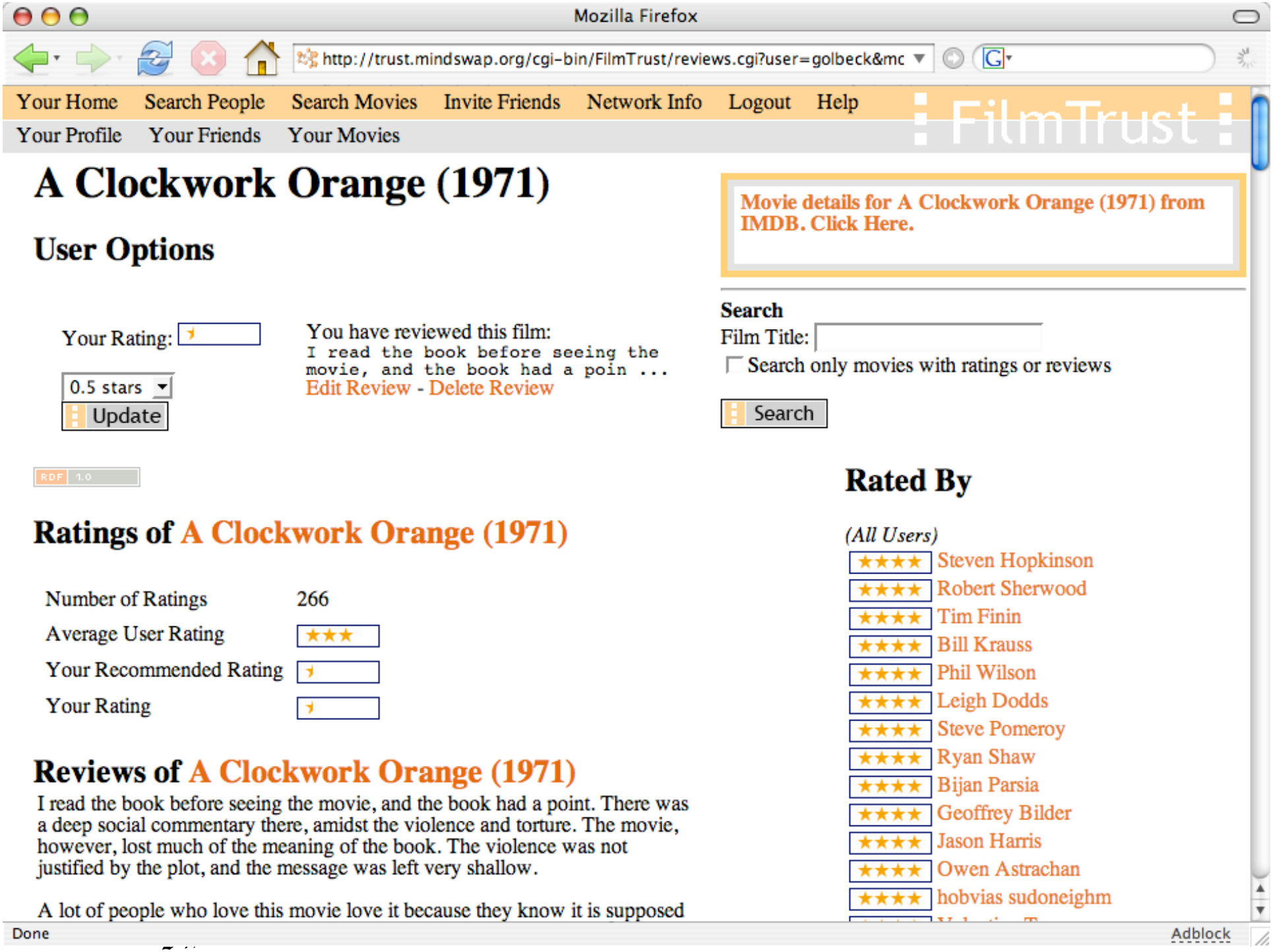
$$t_{is} = \frac{\sum_{j \in adj(j) \mid t_{ij} \geq max} t_{ij} t_{js}}{\sum_{j \in adj(j) \mid t_{ij} \geq max} t_{ij}}$$

FilmTrust

- Users maintain
  - Lists of friends
  - Trust ratings for friends
  - Movie ratings
  - Movie reviews
- http://trust.mindswap.org/FilmTrust


Mozilla Firefox
http://trust.mindswap.org/cgi-bin/FilmTrust/self.cgi?user=test
Your Home
Search People
Search Movies
Invite Friends
Network Info
Logout
Help
Your Profile
Your Friends
Your Movies
FilmTrust
Personal Details
I am the FilmTrust test account - but I have strong opinions and I'm not afraid to use them.
Movie Tidbits
I loved Spice World. It could quite possibly be the best movie ever made.
Personal Options
Update
Test Test
Username: test
FOAF
See this person in the network
FilmTrust, Golbeck 05-06
Done
Adblock


Mozilla Firefox
http://trust.mindswap.org/cgi-bin/FilmTrust/userFriends.cgi?user=go
Your Home
Search People
Search Movies
Invite Friends
Network Info
Logout
Help
Your Profile
Your Friends
Your Movies
FilmTrust
Your Trust Network
Find people:
Everywhere
Search
Jen's Friends
Person
Trust Rating
Options
irene golbeck 10 Update
Test Test 10 Update
Pi Golbeck 10 Update
K Golbeck 10 Update
Dan Z Golbeck 9 Update
Jennifer Boughanem 9 Update
john golbeck 9 Update
paolo massa 8 Update
Sandro Fouche 8 Update
Matthew Velic 8 Update
Jim Hendler 8 Update
Jennifer Sneirson-Kun 7 Update
Bijan Parsia 7 Update
Dave Wang 6 Update
Friends with Jen
Person
Options
Anne-Celine Maurice Add
Aditya Kalyanpur Update
Anant K Add
Stephanie Cross Add
Bijan Parsia Update
Robert Sherwood Add
D Cross Add
Dean Allemang Update
Dan Z Golbeck Update
Evren Sirin Add
Gene Chipman Update
debbie heisler Add
Jennifer Sneirson-Kun Update
Jenneke Fokker Add
Done
Adblock


Mozilla Firefox
http://trust.mindswap.org/cgi-bin/FilmTrust/reviews.cgi?user=golbeck&mc
Your Home
Search People
Search Movies
Invite Friends
Network Info
Logout
Help
Your Profile
Your Friends
Your Movies
FilmTrust
A Clockwork Orange (1971)
User Options
Your Rating:
0.5 stars
Update
You have reviewed this film:
I read the book before seeing the movie, and the book had a poin ...
Edit Review - Delete Review
Movie details for A Clockwork Orange (1971) from IMDB. Click Here.
Search
Film Title:
Search only movies with ratings or reviews
Search
RDF 1.0
Ratings of A Clockwork Orange (1971)
Number of Ratings 266
Average User Rating
Your Recommended Rating
Your Rating
Reviews of A Clockwork Orange (1971)
I read the book before seeing the movie, and the book had a point. There was a deep social commentary there, amidst the violence and torture. The movie, however, lost much of the meaning of the book. The violence was not justified by the plot, and the message was left very shallow.
A lot of people who love this movie love it because they know it is supposed
Rated By
(All Users)
Steven Hopkinson
Robert Sherwood
Tim Finin
Bill Krauss
Phil Wilson
Leigh Dodds
Steve Pomeroy
Ryan Shaw
Bijan Parsia
Geoffrey Bilder
Jason Harris
Owen Astrachan
hobvias sudoneighm
Done
Adblock


Mozilla Firefox
http://trust.mindswap.org/cgi-bin/FilmTrust/reviews.cgi?user=golbeck&mc
Reviews of A Clockwork Orange (1971)
I read the book before seeing the movie, and the book had a point. There was a deep social commentary there, amidst the violence and torture. The movie, however, lost much of the meaning of the book. The violence was not justified by the plot, and the message was left very shallow.
A lot of people who love this movie love it because they know it is supposed to be deep and important, not because there is actually a lot to love here. Read the book instead.
- by Jen Golbeck
This movie sucked! It was probably the worst movie I have ever seen. As I left the theatre, I remember thinking, "what was that all about?"
- by john golbeck
Absolutely the worst movie ever. Very weird, hard to follow and disturbing. Perhaps if the book had been followed better it would have been more tolerable to sit through
- by irene golbeck
One of Stanley Kubricks science fiction classics. Staring Malcolm McDowell as "Alex", a young man whose principle interests are rape, ultra-violence, and Beethoven.
The first half of this film is a disturbing no-holds-barred look at the life of Alex. Kubrick seems to want to make the audience dislike the principle character, with little sympathy for what will ultimately happen to him.
When Alex goes to jail, he appears unrepentent and determined to subvert the system. Instead, he is subjected to a treatment to "cure" him of his ways. After his cure and subsequent release, his past revisits him, causing the audience to
★★★★ Ryan Shaw
★★★★ Bijan Parsia
★★★★ Geoffrey Bilder
★★★★ Jason Harris
★★★★ Owen Astrachan
★★★★ hobvias sudoneighm
★★★★ Valentina Tamma
★★★★ Bernardo Cuenca
★★★★ fox mulder
★★★★ Winnie Kessler
★★★★ Kaan Ege
★★★★ Paulo Pinheiro da Silva
★★★★ Chris Craun
★★★★ Øystein Holm-Olsen
★★★★ Stephen Harris
★★★★ Chris Walton
★★★★ Hermann Keldenich
★★★★ Duy Vu
★★★★ nivas gallo
★★★★ Jenneke Fokker
★★★★ Faith Piper
★★★★ Eric Prud'hommeaux
★★★★ Paul Gearon
★★★★ Michael Rueger
★★★★ Suzan Foster
★★★★ Dietrich Ayala
★★★★ Sean Cier
★★★★ Jochen Notholt
★★★★ michael diamond
★★★★ Davide Eynard
Done
Adblock

tolerable to sit through
- by irene golbeck

★

One of Stanley Kubricks science fiction classics. Staring Malcolm McDowell as "Alex", a young man whose principle interests are rape, ultra-violence, and Beethoven.

The first half of this film is a disturbing no-holds-barred look at the life of Alex. Kubrick seems to want to make the audience dislike the principle character, with little sympathy for what will ultimately happen to him.

When Alex goes to jail, he appears unrepentent and determined to subvert the system. Instead, he is subjected to a treatment to "cure" him of his ways. After his cure and subsequent release, his past revisits him, causing the audience to question the attitudes they have formed about this character.

This is a dark and cynical film about the way society treats its fringe elements. It contains numerous dichotomies, including an incredibly funny sex scene, as well as the most disturbing rape scene I've ever witnessed. The violence is stylized rather than authentic, but Kubrick makes great use of psychological elements with a much greater effect than the highly explicit elements used in more modern films.
- by Paul Gearon

★★★★

Notwithstanding the movie's poor showing in the Golbeck demographic, this is actually an interesting --- if at times difficult to watch --- film. Kubrick does a pretty amazing job of making the audience feel compassion for a sociopath. That said, the film is not 100% faithful to the book. Most notably, the final (redemptive) chapter is omitted. But Kubrick never seemed all that much into redemption anyway, and so the ending is not necessarily what we'd hope for, but since when have Kubrick's films been described as 'life-affirming'? Worth watching, even through hands over your eyes in parts...
- by Brian Shields

★★★★

★★★★ nivas gallo
★★★★ Jenneke Fokker
★★★★ Faith Piper
★★★★ Eric Prud'hommeaux
★★★★ Paul Gearon
★★★★ Michael Rueger
★★★★ Suzan Foster
★★★★ Dietrich Ayala
★★★★ Sean Cier
★★★★ Jochen Notholt
★★★★ michael diamond
★★★★ Davide Eynard
★★★★ Lorie Huertas
★★★★ karsten schmidt
★★★★ Sebastian Rodriguez
★★★★ theo takiari
★★★★ Mónica Rojas
★★★★ Brad Bebee
★★★★ Santiago Molina
★★★★ Rachel Rein
★★★★ Al Gray
★★★★ Maria Leonard
★★★★ jennifer lee
★★★★ Darin Marshall
★★★★ Eric Miller
★★★★ Bill McDaniel
★★★★ Ron Craswell
★★★★ Greg Tyrelle
★★★★ Caz Ksiazek
★★★★ John Smart
★★★★ Jim McManus
★★★★ Lyndsay Burtonshaw
★★★★ Simon Jacquier

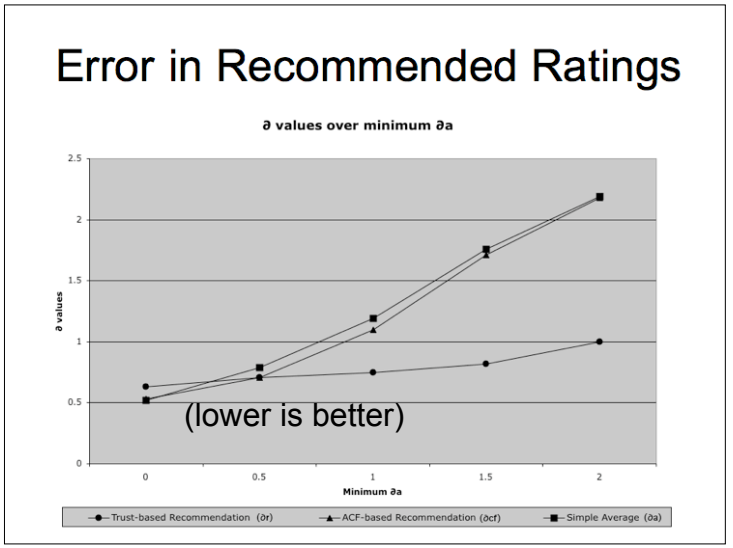
Trust-enhanced browsing

Bringing the social into the Web interface

Rensselaer

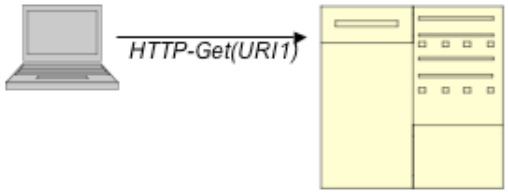
Challenges

- Can outperform recommender systems
  - Especially where the individual's rating is further from the group's
- Continues the Web's challenge on the concept of "ground truth"
  - What happens when this is applied to more complex Web based commerce than recommenders?

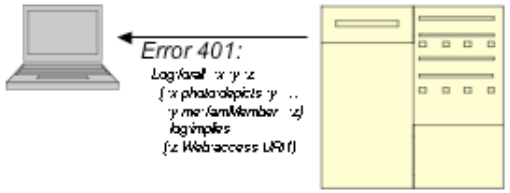


Rensselaer

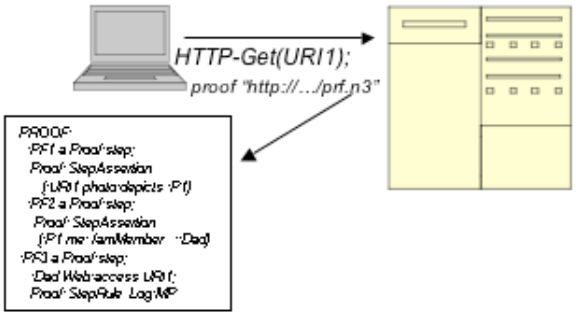
Example 2: Policy Aware Web



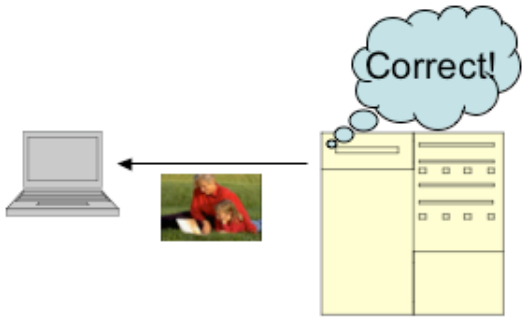
(A) User requests a resource.



(B) 401 error provides access rules.



(C) Proof is generated and pointer is sent in new HTTP-Get request.



(D) Proof is checked, and confirmed, and the transaction succeeds.

(NSF ITR; Hendler, Berners-Lee, Weitzner; 2005)

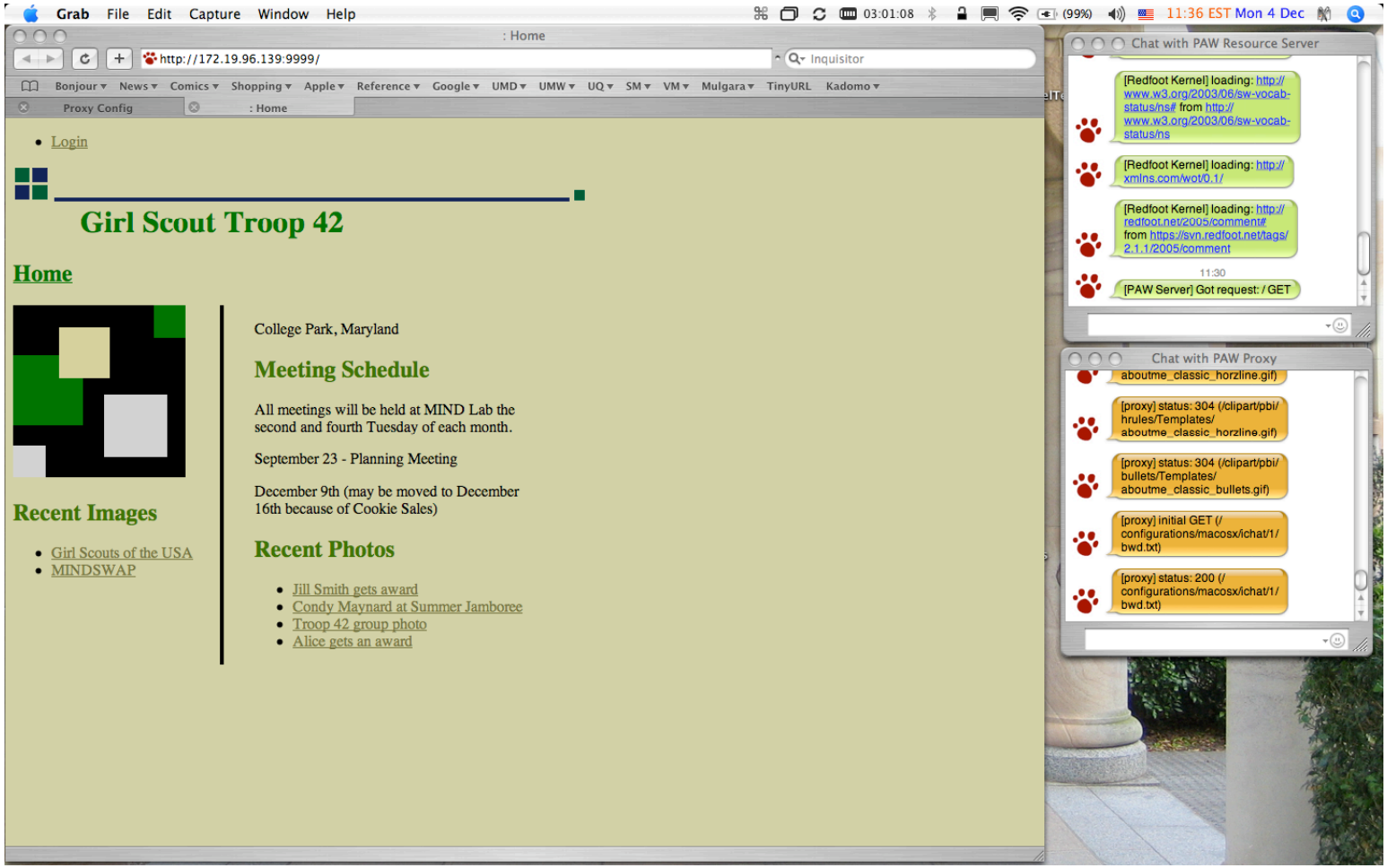
Rensselaer


Girl Scout Troup 42: Home Page
http://demo.policyawareweb.org/
Inquisitor
Girl Scout Troup 42
Home
College Park, Maryland
Meeting Schedule
All meetings will be held at MIND Lab the second and fourth Tuesday of each month.
September 23 - Planning Meeting
December 9th (may be moved to December 16th because of Cookie Sales)
Recent Images
Girl Scouts of the USA
MINDSWAP
Recent Photos
Jill Smith gets award
Condy Maynard at Summer Jamboree
Troop 42 group photo

Rensselaer

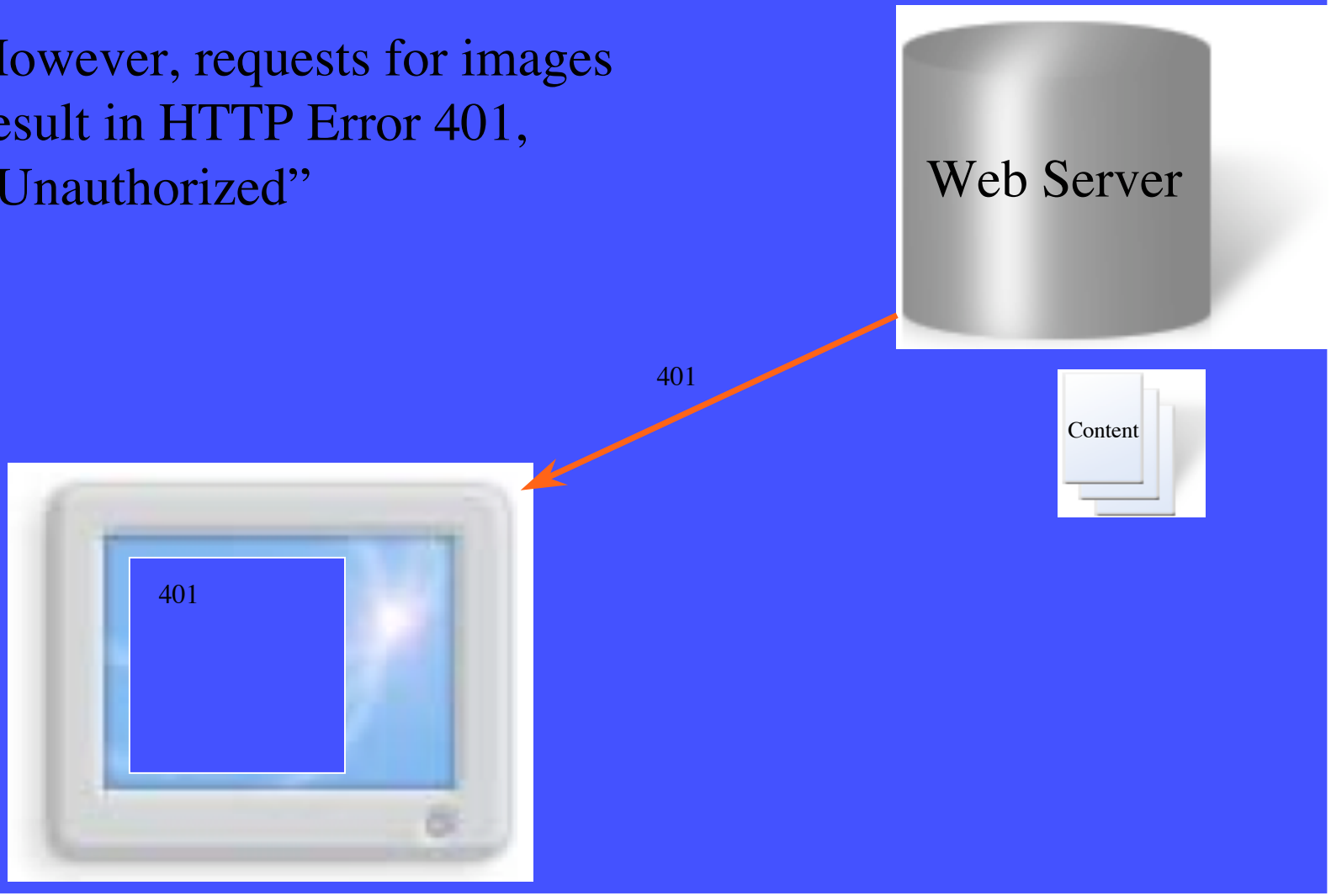
Use case:
A Web browser requests the home page for a girl scout troop and is given it by a Web server.

Rensselaer


Grab File Edit Capture Window Help
03:01:08
11:36 EST Mon 4 Dec
: Home
http://172.19.96.139:9999/
Inquisitor
Bonjour News Comics Shopping Apple Reference Google UMD UMW UQ SM VM Mulgara TinyURL Kadomo
Proxy Config
: Home
Login
Girl Scout Troop 42
Home
College Park, Maryland
Meeting Schedule
All meetings will be held at MIND Lab the second and fourth Tuesday of each month.
September 23 - Planning Meeting
December 9th (may be moved to December 16th because of Cookie Sales)
Recent Images
Girl Scouts of the USA
MINDSWAP
Recent Photos
Jill Smith gets award
Condy Maynard at Summer Jamboree
Troop 42 group photo
Alice gets an award
Chat with PAW Resource Server
[Redfoot Kernel] loading: http://www.w3.org/2003/06/sw-vocab-status/ns# from http://www.w3.org/2003/06/sw-vocab-status/ns
[Redfoot Kernel] loading: http://xmlns.com/wot/0.1/
[Redfoot Kernel] loading: http://redfoot.net/2005/comment# from https://svn.redfoot.net/tags/2.1.1/2005/comment
11:30
[PAW Server] Got request: / GET
Chat with PAW Proxy
aboutme_classic_horzline.gif)
[proxy] status: 304 (/clipart/pbi/hrules/Templates/aboutme_classic_horzline.gif)
[proxy] status: 304 (/clipart/pbi/bullets/Templates/aboutme_classic_bullets.gif)
[proxy] initial GET (/configurations/macosx/ichat/1/bwd.txt)
[proxy] status: 200 (/configurations/macosx/ichat/1/bwd.txt)

Rensselaer

However, requests for images result in HTTP Error 401, "Unauthorized"



Rensselaer

The 401 “Unauthorized” response has been modified to provide a URL to a policy:

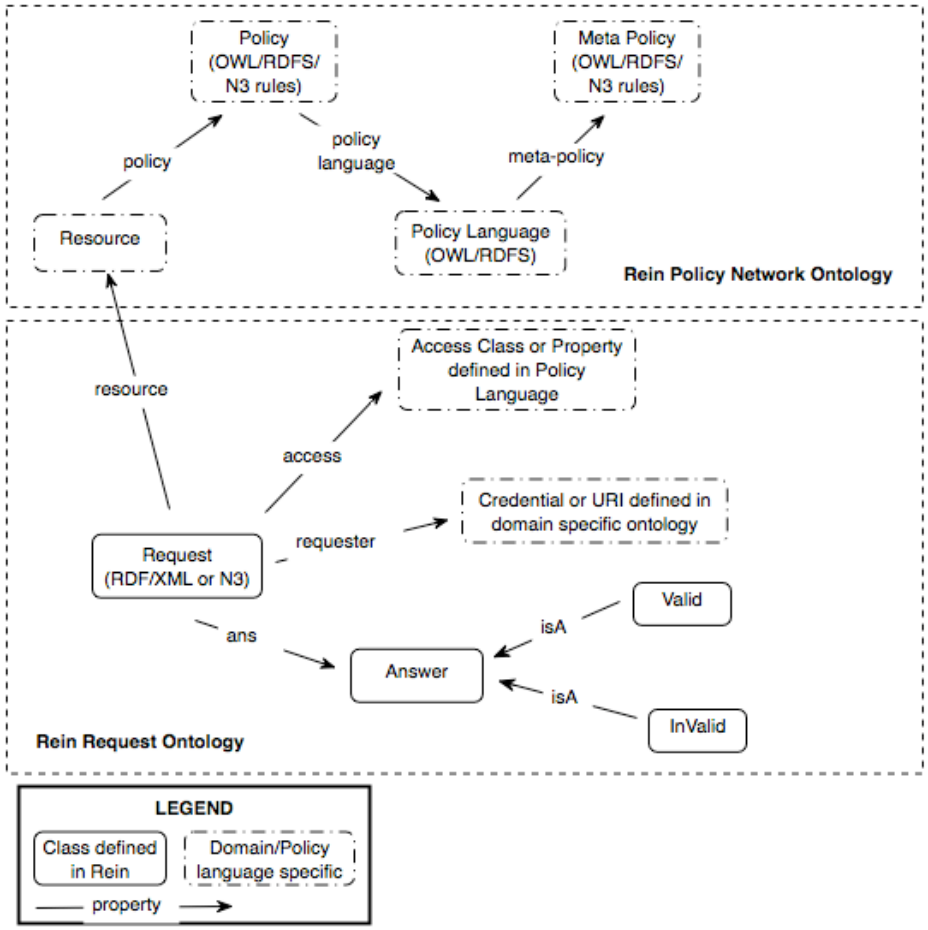
```
HTTP/1.1 401 Not authorized
 Date: Sat, 03 Dec 2005 15:32:18 GMT
 Server: TwistedWeb/2.0.1
 Policy: http://groups.csail.mit.edu/dig/2005/09/rein/examples/troop42-policy.n3
 Content-type: text/html; charset=UTF-8
 Connection: close
10:32:20 ERROR 401: Not authorized.
```

Policies use distributed rules

- Example policies
  - Photos taken at meetings of the troop can be shared with any current member of the troop.
  - Photos taken at a jamboree can be shared with anyone in the troop or with anyone who attended the jamboree.
  - Photos of any girl in the troop can be shared with the world if that girl's parent has given permission

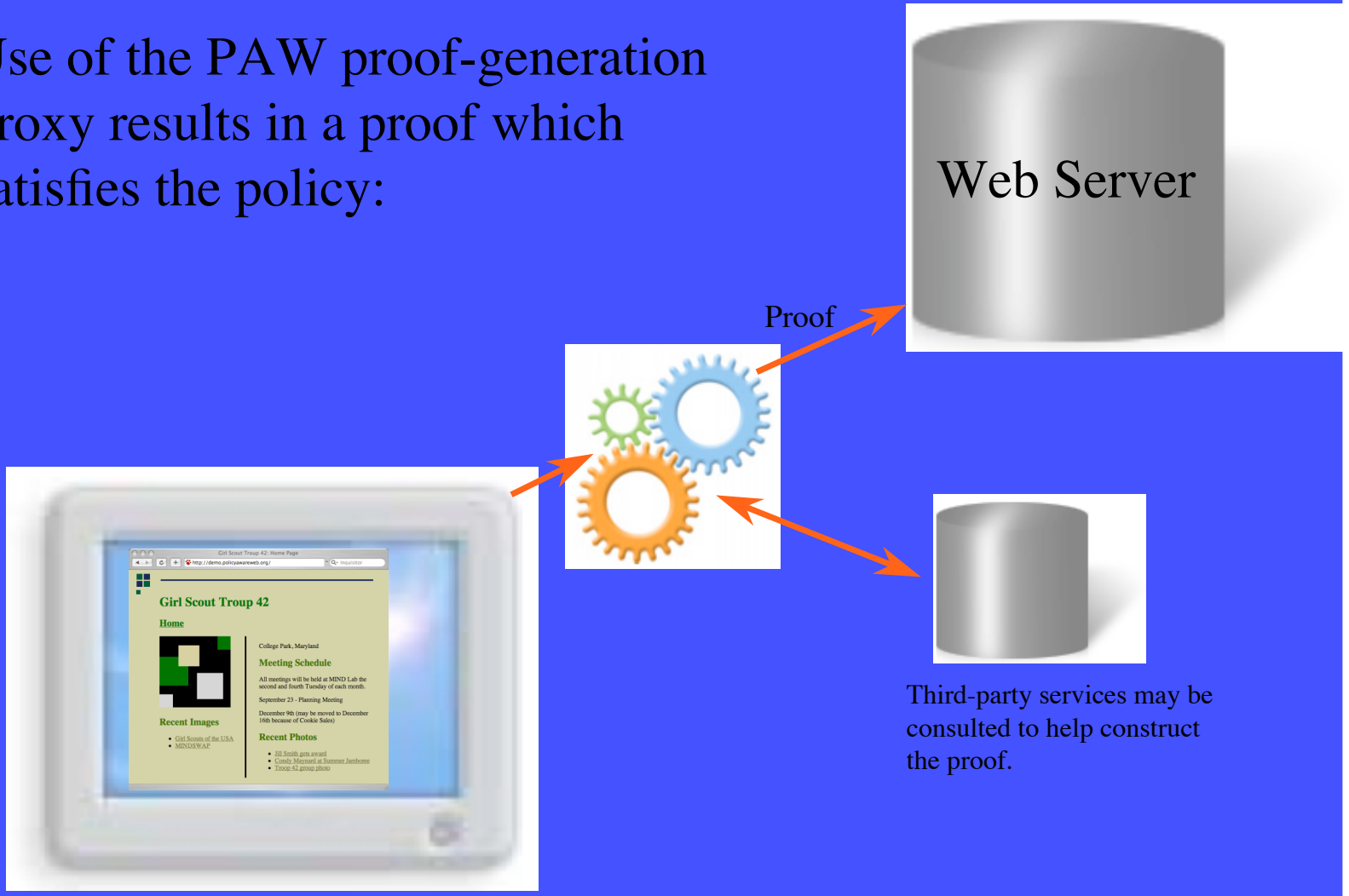
```
{ REQ a rein:Request.
  REQ rein:resource PHOTO.
  ?F a TroopStuff; log:includes
       { PHOTO a t:Photo; t:location LOC.
         LOC a t:Meeting }.

  REQ rein:requester WHO.
  WHO session:secret ?S.
  ?S crypto:md5 TXT.

  ?F a TroopStuff; log:includes
       { [] t:member [ is foaf:maker of PG ].
         LOC t:attendee [ is foaf:maker of PG ] }.
  PG log:semantics [ log:includes
       { PG foaf:maker [ session:hexdigest TXT ] }
   ].

} => { WHO  http:can-get PHOTO }.
```

Rein "ontology"



Rensselaer

Use of the PAW proof-generation proxy results in a proof which satisfies the policy:

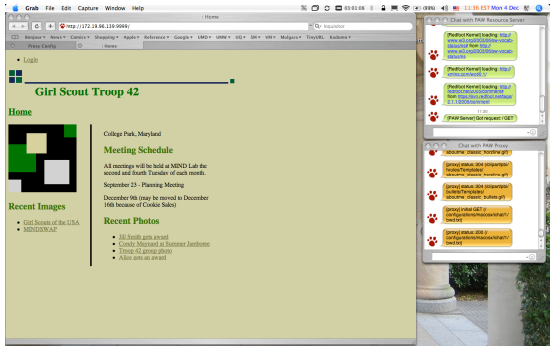


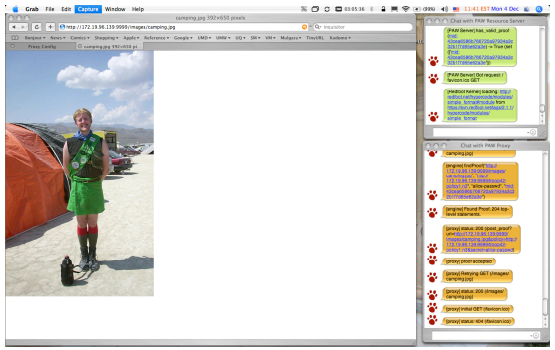
Rensselaer

The Web server checks the proof and serves the content if it is valid.

Challenges

- Allows more control over distribution of personal information
  - Although we are not yet there
- Increases the Web's tension between privacy and "desire to know"
  - How do we allow individuals to easily control their identities without revealing too much through the aggregation of this information





Rensselaer

And more to come…

- Consider people having multiple identities and multiple attributes in workspaces (not just play) which include virtual as well as real identities

AIAI2 – Second Life
secondlife://Saengseon/197/42/108

Conclusion

- The Web is a complex and messy place
  - Some "order" added by Semantic Web, but growing in ways not being explored by many AI researchers
- The Web is evolving in many complex ways
  - Today's example, social issues in Web use
    - New functionalities
    - But potentially disruptive technologies
- This drives us towards a new agenda
  - Understanding the Web in a Scientific Way
    - Modeling, engineering, and especially, social impact

Learn more at http://www.webscience.org
(Web Science Research Initiative)

Tetherless ~~World~~ Web

- The pervasive, mobile and ubiquitous Web, requires:
  - Understanding the Web as a network at all scales
    - From protocol to social network levels
    - Non-browser access issues (above the IP level)
      - Finer-grained information access
  - New methods for
    - Privacy protection
    - Security needs
  - Policy Aware Computing
    - Context dependent information use

But that's a talk for some future date…