

LAST WEEK ON IO LAB



Project 2 was due today at noon. If you haven't sent it to us and haven't already talked to us, come talk to us now.

Be ready to demo and discuss in class.



We learned about **version control** and **Subversion**.



INFORMATION ORGANIZATION LAB

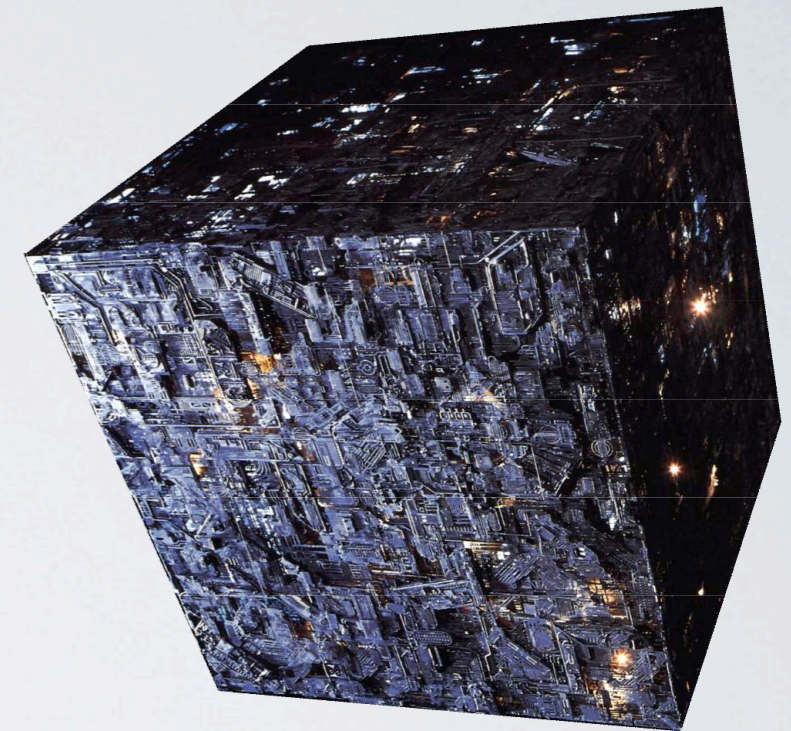
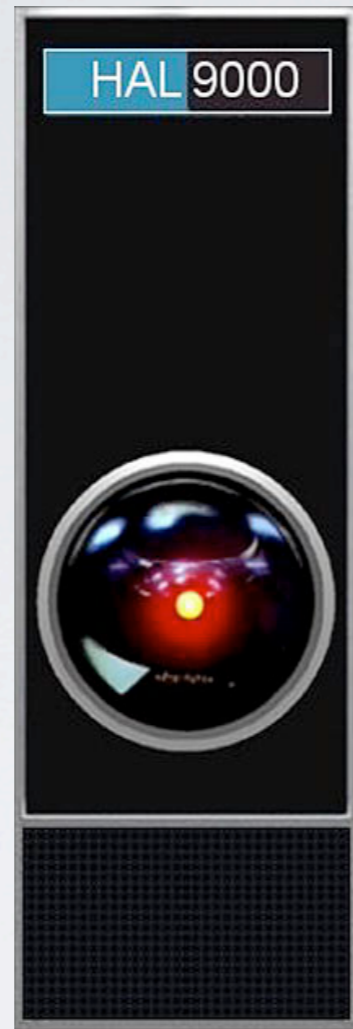
PROJECT 2



The New York Times

Demo rules:

1. Everybody has to tell us something. (doesn't have to be a dissertation, though)
2. If it doesn't work, just explain what is supposed to happen and why you did it.
3. Remember to zoom in.



THE SEMANTIC WEB

Information made for machines

The semantic web means a bunch of different things to a different people. Remember reading three articles in 202 from the meta-utopia of an agent browsing the web on your behalf, to ever-so-slightly smarter applications of semantic information.



microformats

<http://microformats.org>

Semantic Web, Part 1: Microformats. Small bits of semantic information. Agreed upon standards for adding machine readable data to HTML pages. One of the philosophies behind MF: machine data that nobody sees doesn't get properly maintained. We should apply machine identifiers to existing visible data.



microformats

RelLicense

hCard

hCalendar

hReview

Microformats add semantic, machine-readable information to existing content. Semantic because it has an agreed upon meaning. Machine readable because computers can easily retrieve the information (IO/IR tradeoff).

Is anybody using microformats? The microformats website gives examples of each “in the wild”: <http://microformats.org/wiki/hcard-examples-in-wild>. In the case of many formats, usage is somewhat limited.



RelLicense

Demo at <http://creativecommons.org>, <http://blog.stackoverflow.com>.



microformats

hCard

<http://microformats.org/code/hcard/creator>

or

<http://bit.ly/hcardcreator>



RDF & OWL

There are a couple of downsides to Microformats. First, a community has to agree on a standard, which can take time. And then many of these microformats are special cases that require special-case code.

The solution is more complex than microformats, and it's something that you may remember from recent (or not so recent) 202 classes: RDF, OWL, and triples.

TRIPLES

Subject-predicate-object

Andrew Stanton directed Finding Nemo

http://www.freebase.com/view/en/andrew_stanton

<http://ischool.berkeley.edu/verbs/directed>

http://www.freebase.com/view/en/finding_nemo

Last week when Nick was talking about controlled vocabularies, he discussed the idea of URIs as a solution to vocabulary problems.

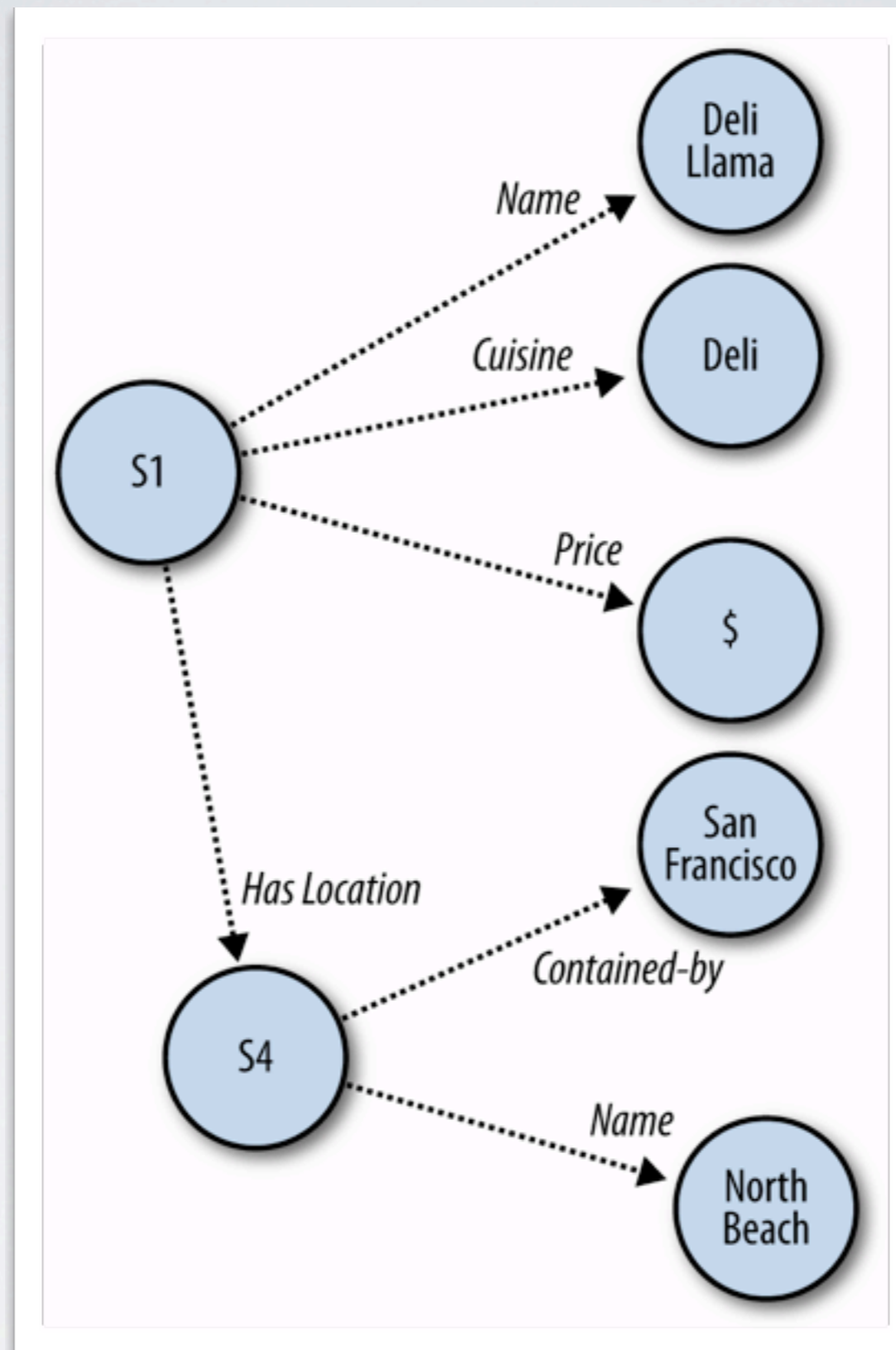
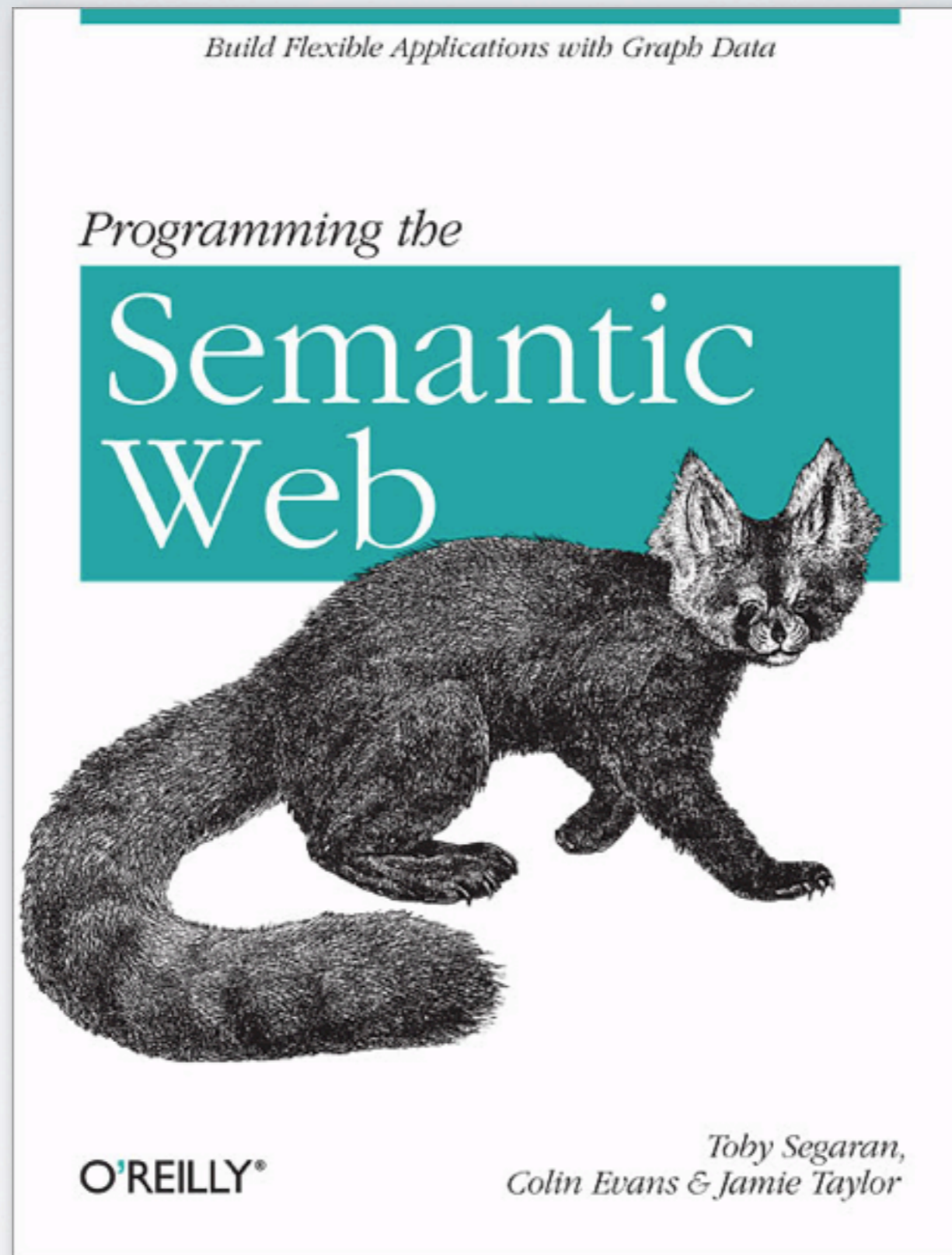


Illustration from *Programming the Semantic Web*, Figure 2.2

Using these triples, you can generate graphs of information about entities that computers can parse and understand. Figure 2-2. A graph of triples showing information about a restaurant. From Prog. sem web chapter 2.



We've assigned some reading from this book, Prog. Sem. Web for next week to augment the areas where we don't know as much. Available from O'Reilly Safari online for free. Talk about type of reading to be done (skim, don't worry too much about the specific implementation in book).

You can read this book while at UC Berkeley at <http://proquest.safaribooksonline.com/9780596802141/>



GOOGLE APP ENGINE

Python + Google = Rapid Prototype Dream

Demo time!

FOR NEXT WEEK



Form groups for project 3.



**Read Chapters 2 & 3 and
Skim Chapter 4 & 5 of
*Programming the Semantic Web.***

Programming the Semantic Web is available at <http://proquest.safaribooksonline.com/9780596802141>.