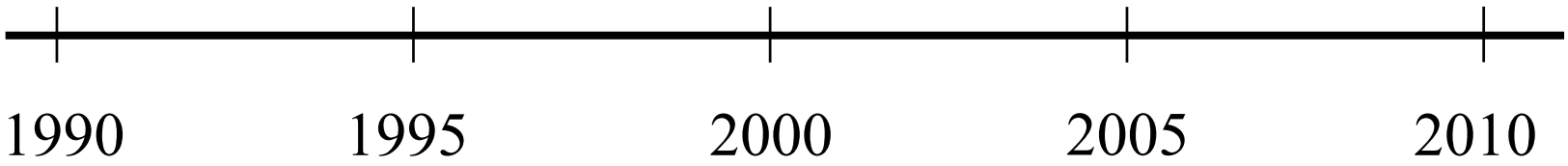
Web Information Management: Past, Present and Future

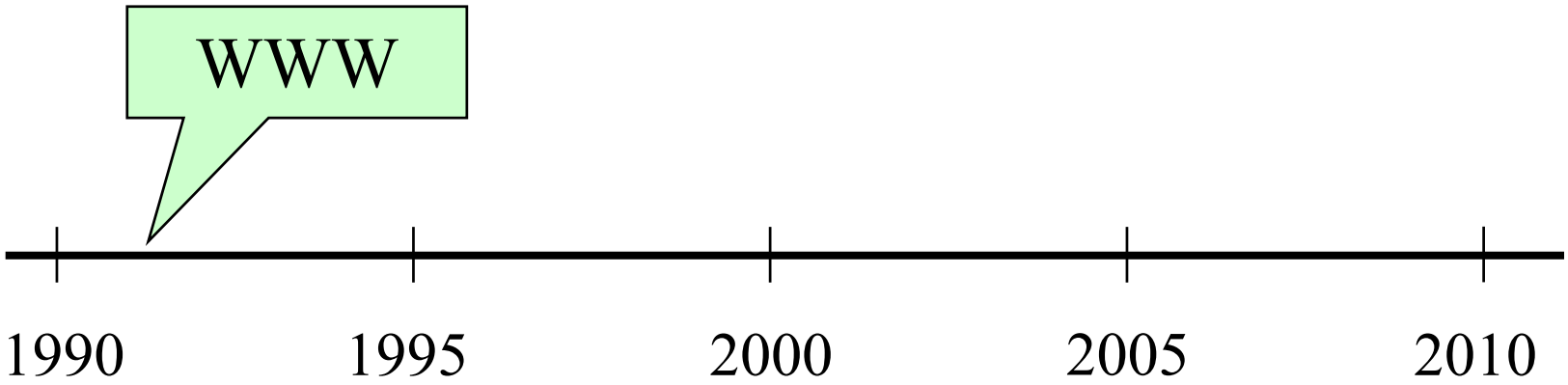
Hector Garcia-Molina

Stanford University

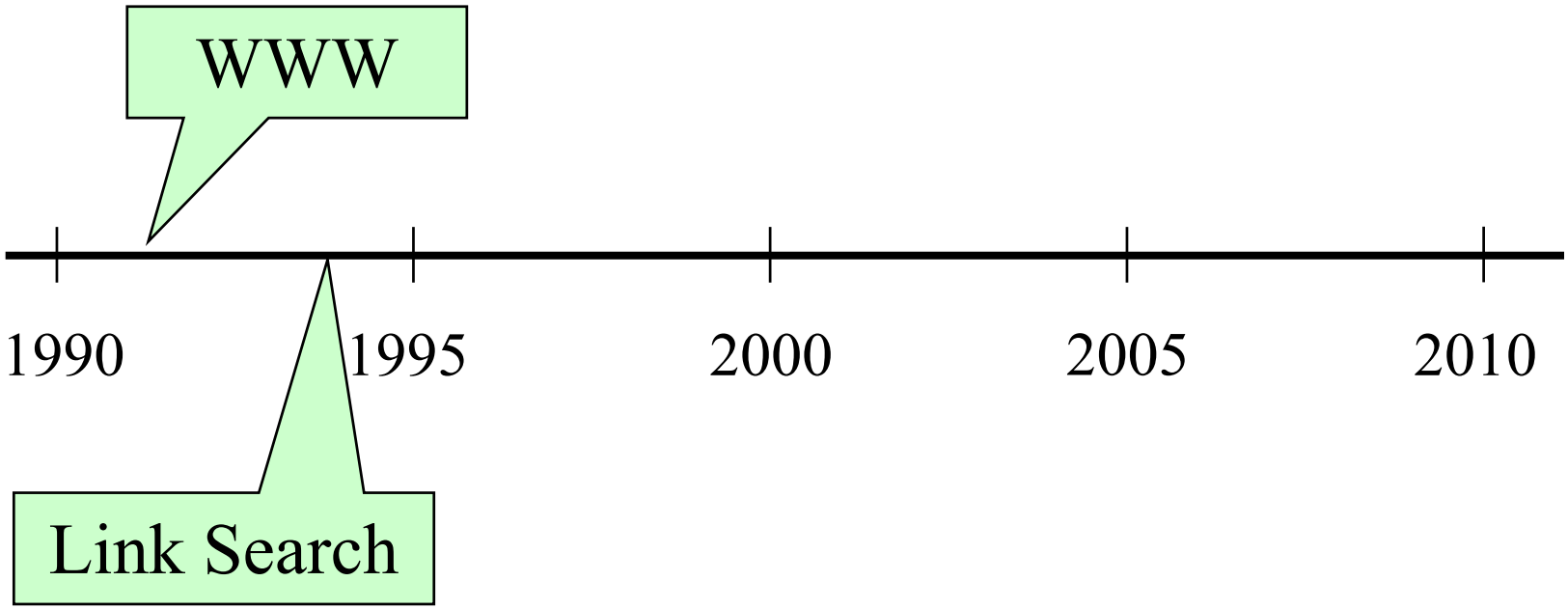
Hector's Web Timeline



Hector's Web Timeline

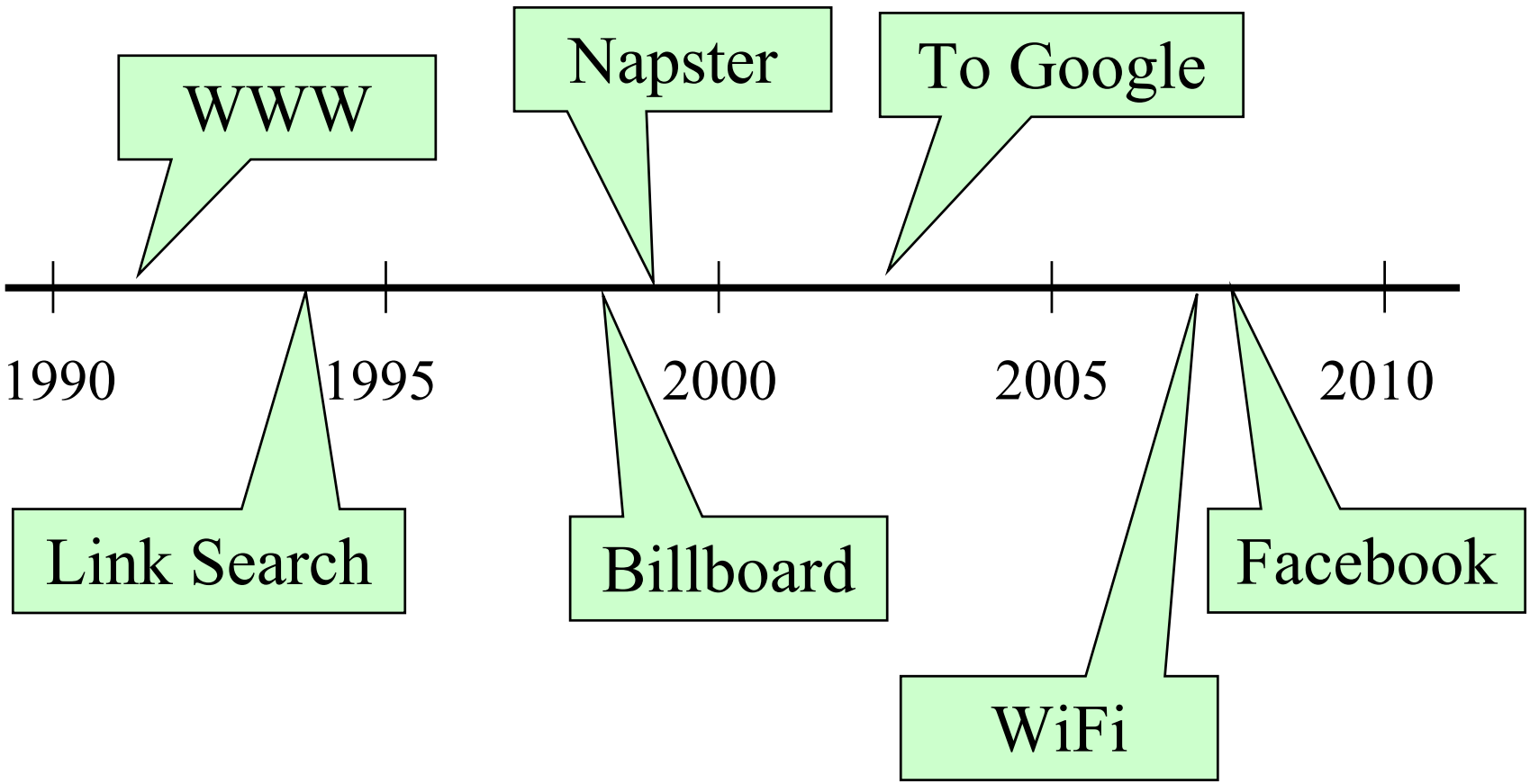


Hector's Web Timeline



Hector Advising

Hector's Web Timeline



rcia - Microsoft Internet Explorer

es Tools Help

Search Favorites

book.com/profile.php?id=7302746

facebook

Profile edit Friends Networks Inbox home account privacy logout

Search

Applications edit

Photos
Groups
Events
Marketplace
more

See Who's Looking for You

Have you ever wondered who is searching for you? Find out instantly at SearchingForMe.com

More Ads | Advertise

View Photos of David (197)
Send David a Gift
Send David a Message
Poke Him!

Mutual Friends

1 friend in common. See All

David Garcia

Networks: Santa Clara Alum, Stanford Alum, National Aeronautics and Space Administration (NASA), Silicon Valley, CA
Sex: Male
Interested In: Women
Relationship Status: In a Relationship with Megan Sabo (Santa Clara)
Birthday: March 2, 1983
Hometown: Monterrey, Mexico
Political Views: Other

Mini-Feed

Displaying 6 stories. See All

January 25
David poops.

January 18
David edited Interests in his profile.

January 17
David wrote on the wall for the event Blow up.

January 15
David plans to attend Blow up.

January 10
David drew a Graffiti for Arturo Hernandez Leal.

More graffiti by David | Draw for David | Draw for Arturo

Marcela Morales tagged David in a photo.

Tagged in: home

Information

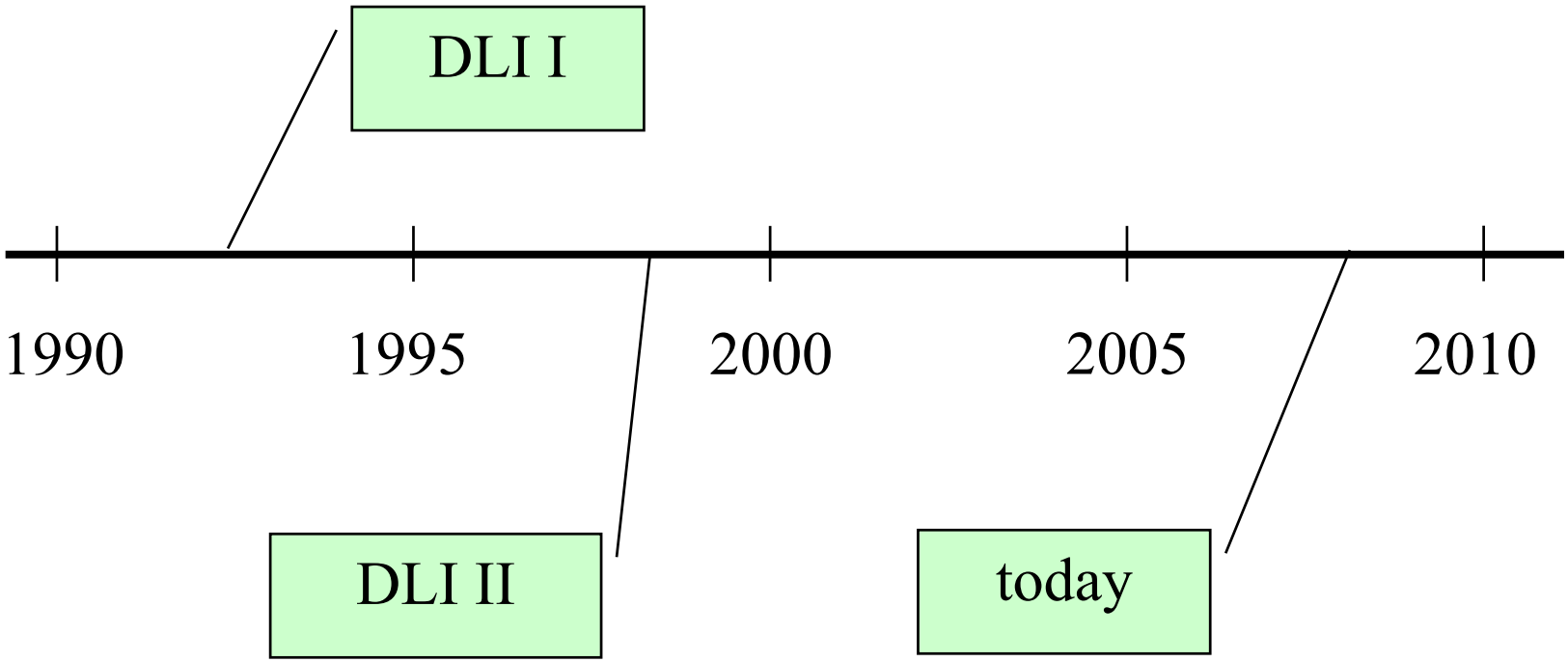
Contact Info

Emails: david.agg@gmail.com, david@db.stanford.edu, dgarcia@email.arc.nasa.gov, dgarcia@scu.edu
Windows Live: X_I_X@hotmail.com
Current Address: 1325 Alviso
Residence: The Den

Personal Info

Activities: 1) drinking 2) writing bestselling literary fiction 3) coding part time for the big bucks
Interests: photography, cooking, novels, chess, videogames, anime, computers, scuba diving

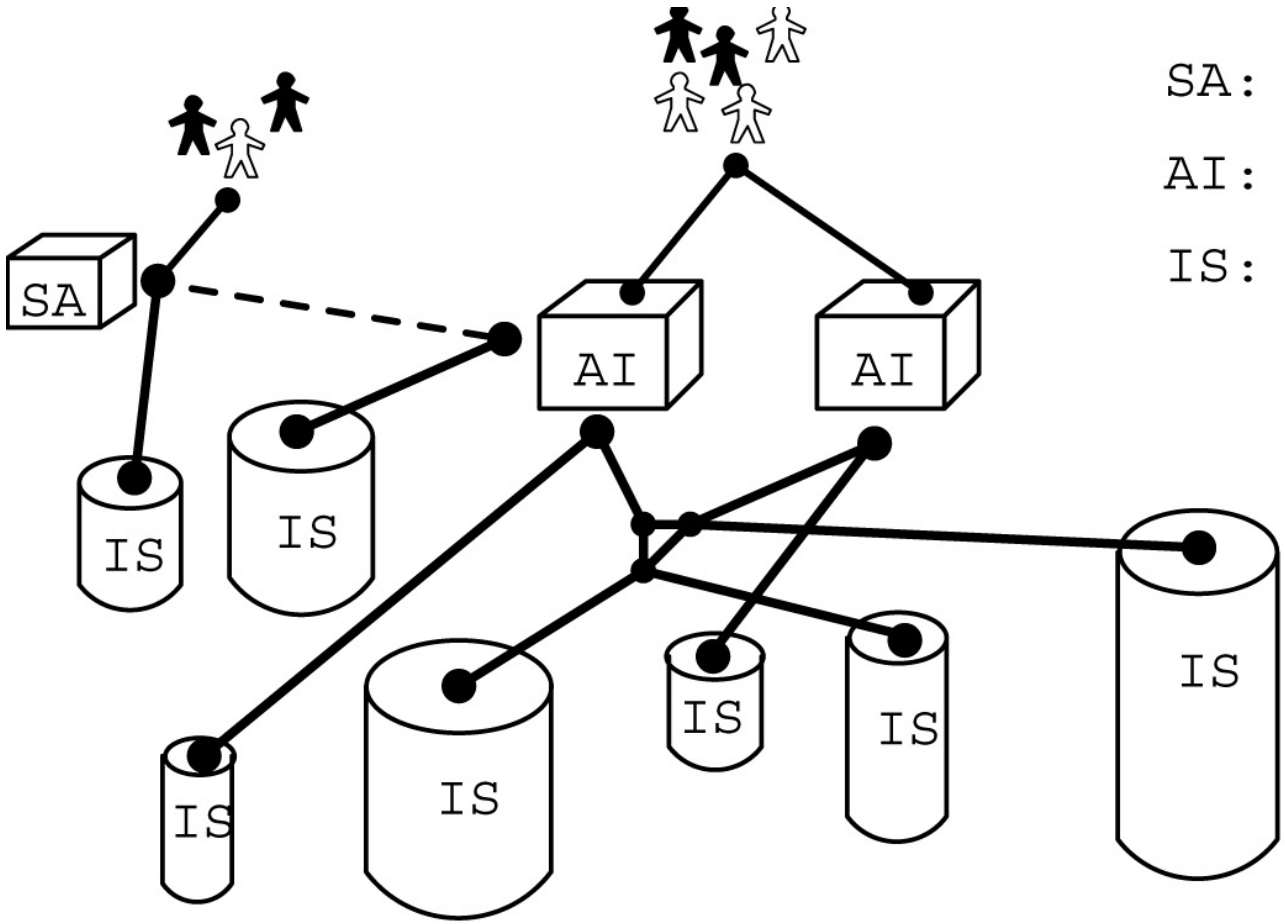
Hector's Challenges



Goal of SDILIP

- The goal of the Stanford Integrated Digital Library Project (SIDLP) is to develop the enabling technologies for a single, integrated and "universal" library, composed from the large numbers of emerging individual heterogeneous repositories.
- Our definition of a constituent repository is broad enough to include everything from personal information collections (e.g., a file of electronic messages) to the collections that one finds today in conventional libraries to the large data collections shared by scientists.

Digital Libraries in 1993

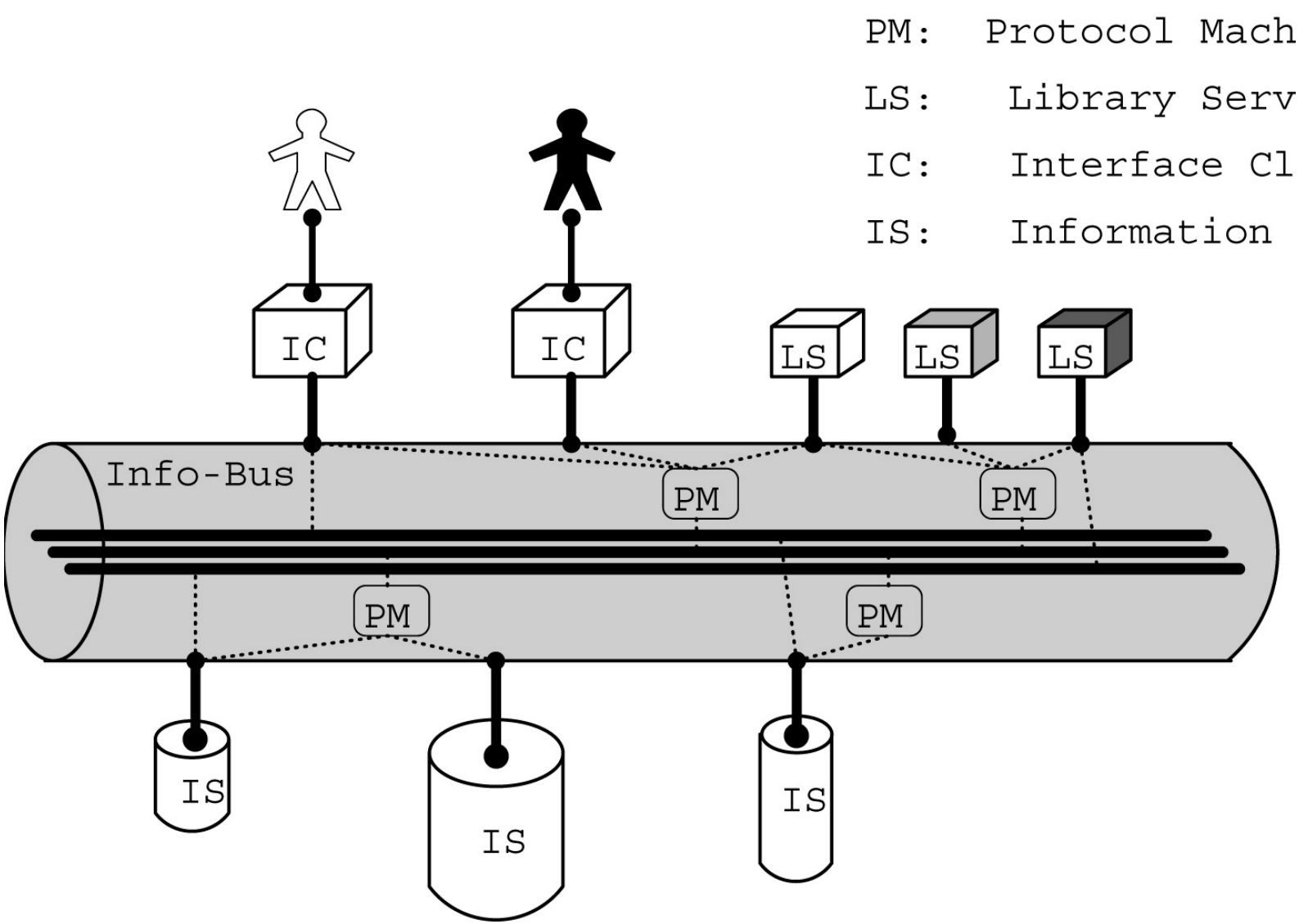


SA: Service Appli

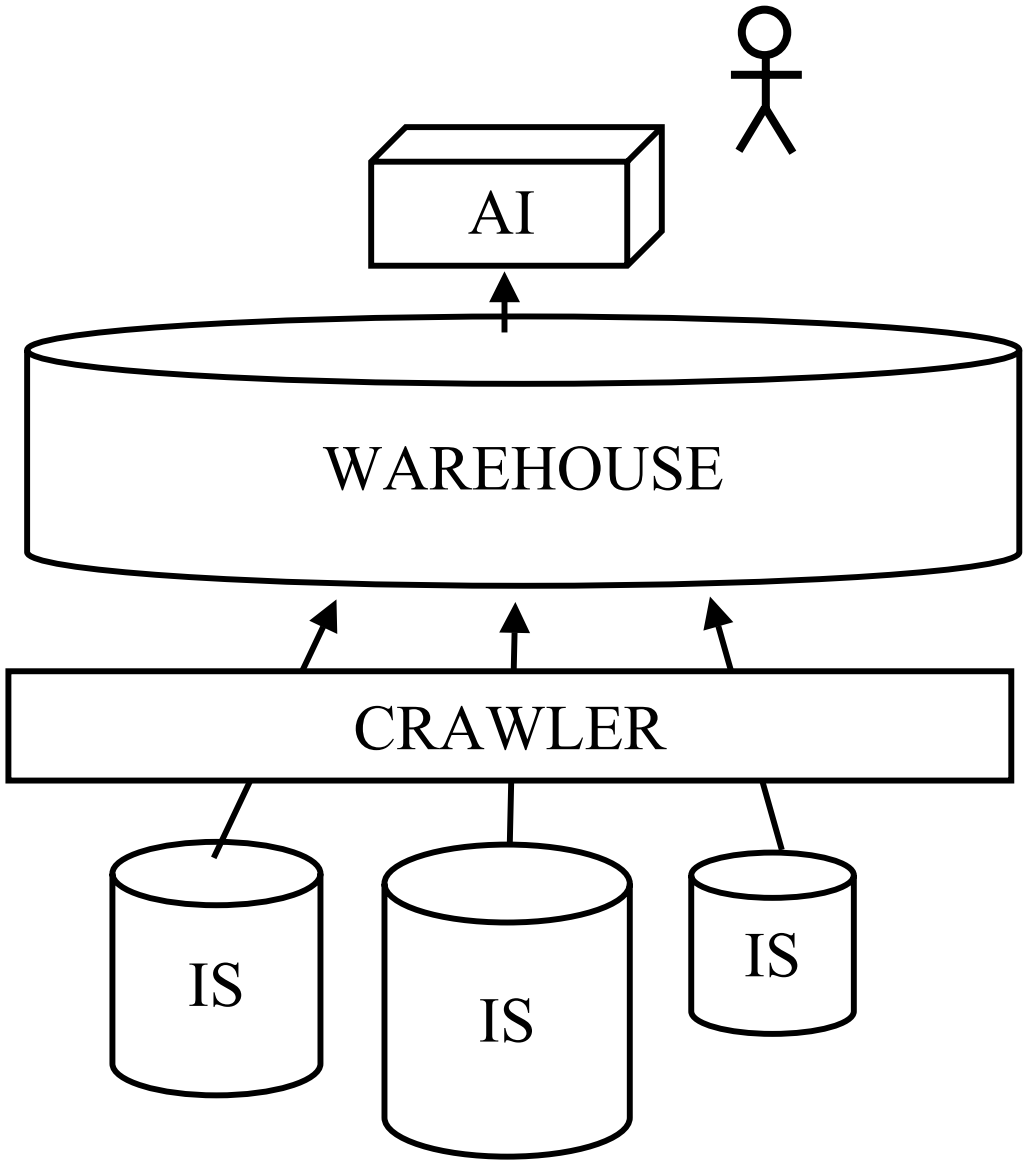
AI: Access Inter

IS: Information

Proposed InfoBus



Alternative to InfoBus



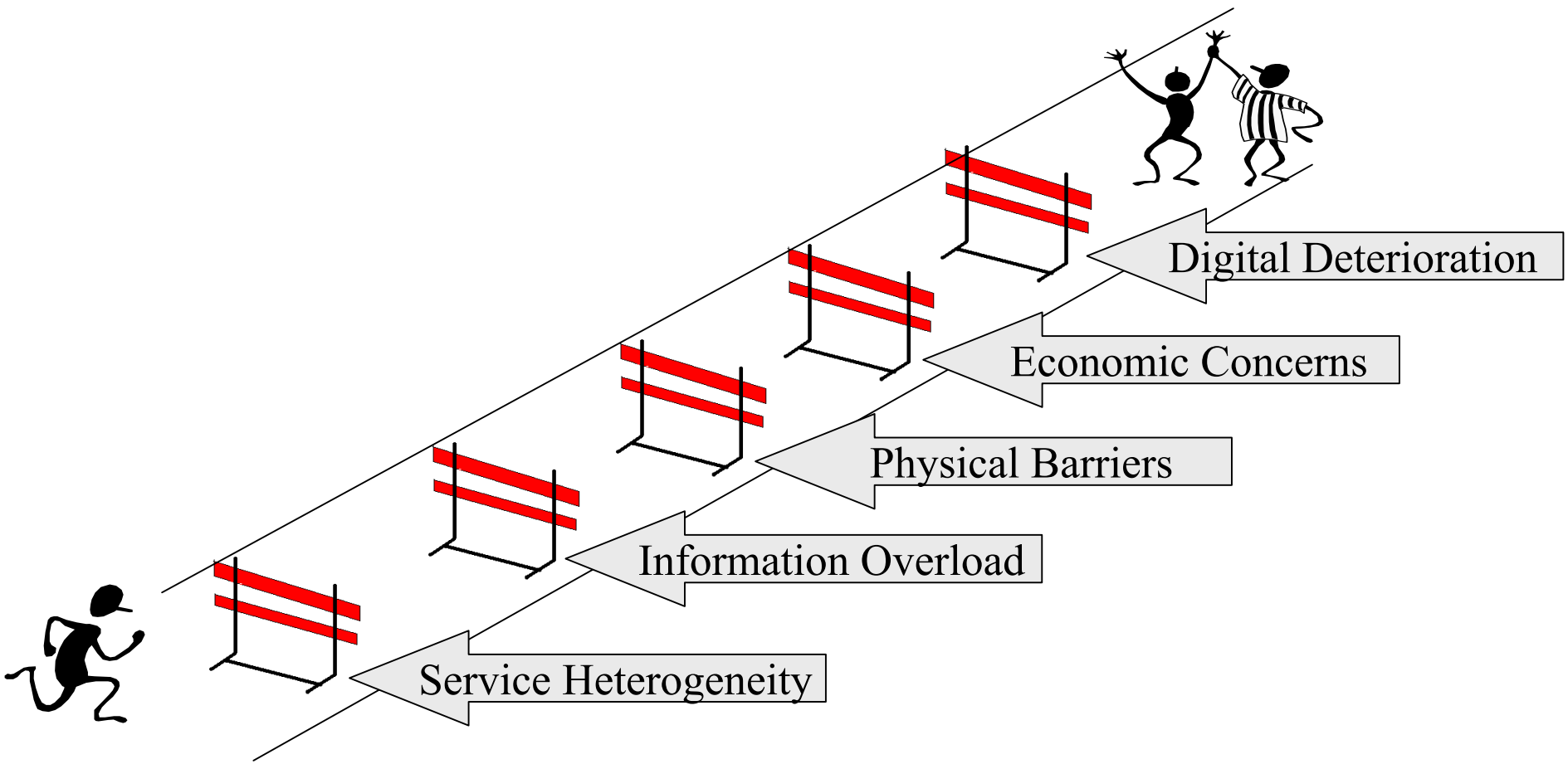
Challenging Problems (1993) Part 1

- What is the proper architecture for the Digital Library?
- What are appropriate models for documents (and more generally information) and library services?
- What protocols for interoperation between repositories, libraries, and services are required?
- What are the best human interfaces to such an integrated library?

Challenging Problems (1993) Part 2

- Making the Digital Library economically viable?
- How does one locate information of interest, in a very large, distributed, and possibly disconnected collection of repositories?
- How does one represent and manipulate the information processing "activities" that occur or will occur in a Digital Library?

Challenging Problems (1998)



Challenging Problems (1998) Part 1

- Service Heterogeneity
  - Different formats, protocols
- Information Overload
  - User's search may return huge numbers of documents
- Physical Barriers
  - User may not be at his desk, or may be connected with insufficient bandwidth and local processing capacity

Challenging Problems (1998) Part 2

- Economic Concerns
  - Document not available due to security or intellectual property protection concerns
- Digital Deterioration
  - The document may have been available in the past, but is now lost.

Challenging Problems 2008 (Part 1)

- Beyond Search
  - identifying user task/intention
  - document/word semantics
- Information Integration
  - extraction, entity resolution
  - combining results
- Monetizing
  - ads, bids,...
  - spam, click fraud

Challenging Problems 2008 (Part 2)

- Social Networks
  - modeling
  - wisdom of the crowds
- Data Mining
  - media mining
  - mining graphs
- Privacy
  - safe data mining
  - protecting identity

Challenging Problems 2008 (Part 3)

- Coping with Scale
  - power minimization
  - revisiting distributed databases
- Personalization
  - access to personal data
  - tailoring services to me
- Mobile Access
  - small devices
  - peer-to-peer libraries

Ranking Challenges

Highest Impact (If Solved)

 # 1 • Beyond Search

 # 2 • Information Integration

 # 3 • Monetizing

 # 4 • Social Networks

- Data Mining ← **low**
- Privacy ← **low**

 # 5 • Coping with Scale

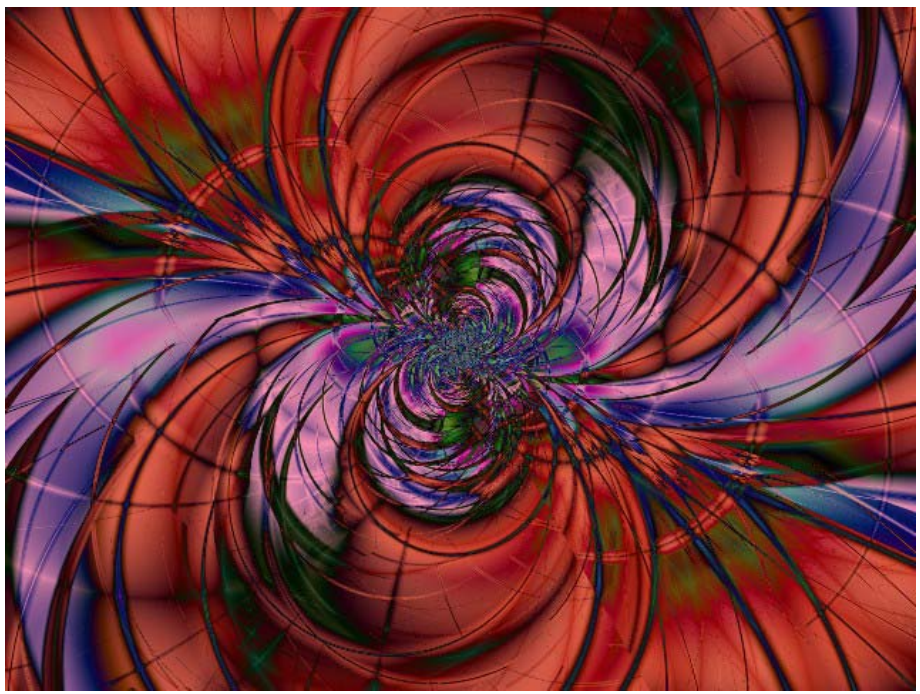
- Personalization ← **low**
- Mobile Access ← **low**

Hardest Problems to Solve

- **# 2** Beyond Search
- **# 1** Information Integration
- **# 3** Monetizing
- **# 4** Social Networks
- Data Mining ← **easy**
- **# 5** Privacy
- Coping with Scale ← **easy**
- Personalization ← **easy**
- Mobile Access ← **easy**

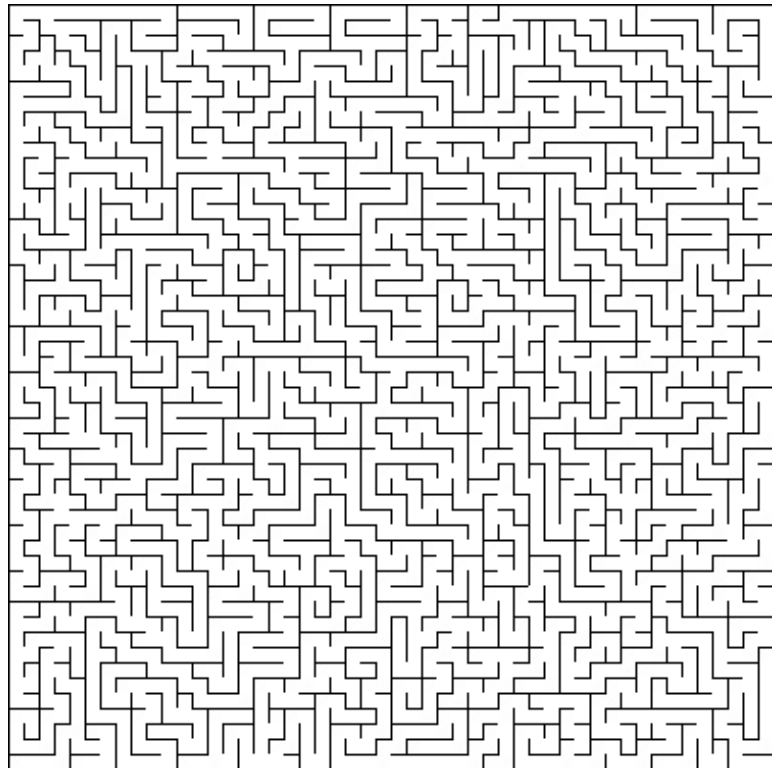
Random Thought 1: Convergence

- DBMSs have IR facilities
- search engines building own DBMS
  - e.g., map-reduce, big table



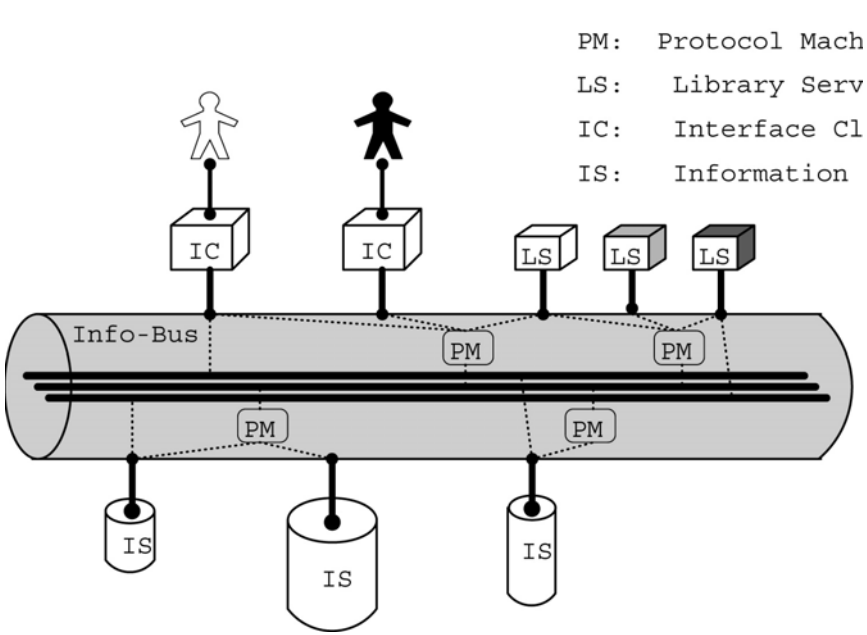
Random Thought 2: Getting Harder

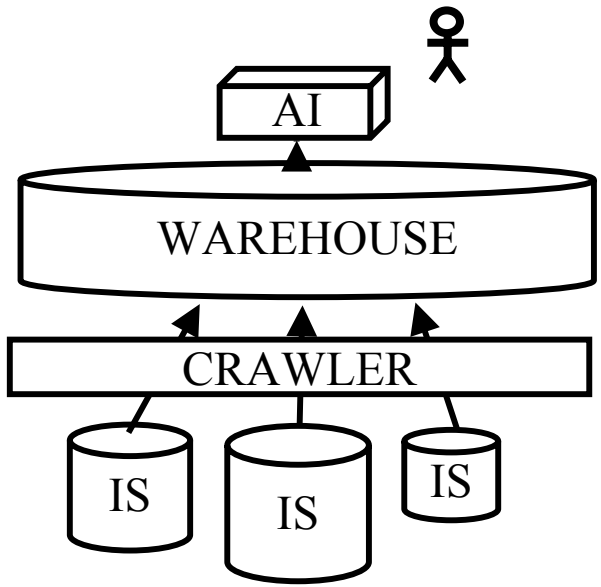
- Web pages more complex
  - less text, flash, ajax, jpgs, deep web,...
  - harder to crawl, index, interpret
- Low Hanging Fruit Taken
  - harder to improve precision, recall
  - more constraints: privacy, mobility,...



Though 3: Back To The Future?

- Is it time to revisit the InfoBus?

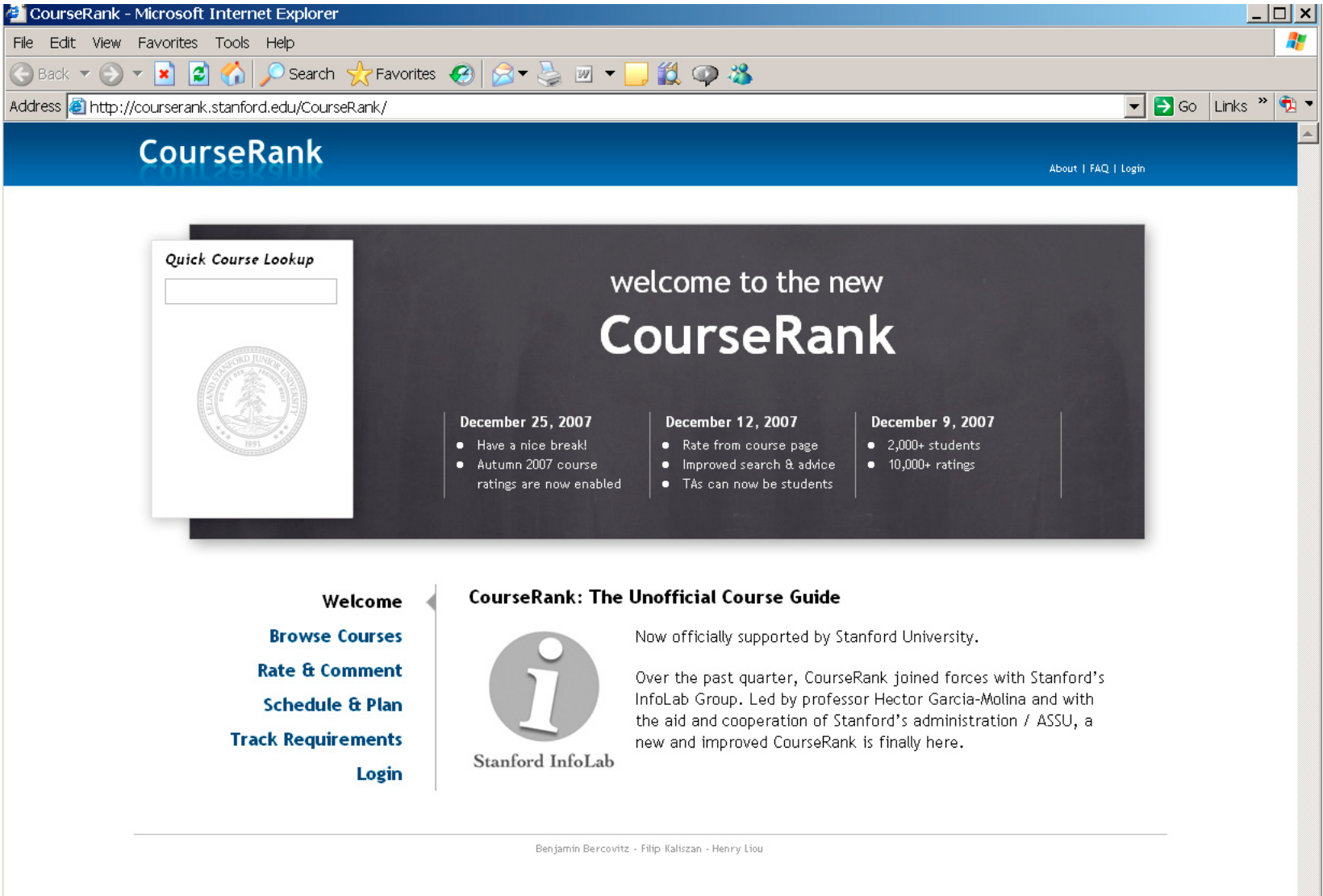


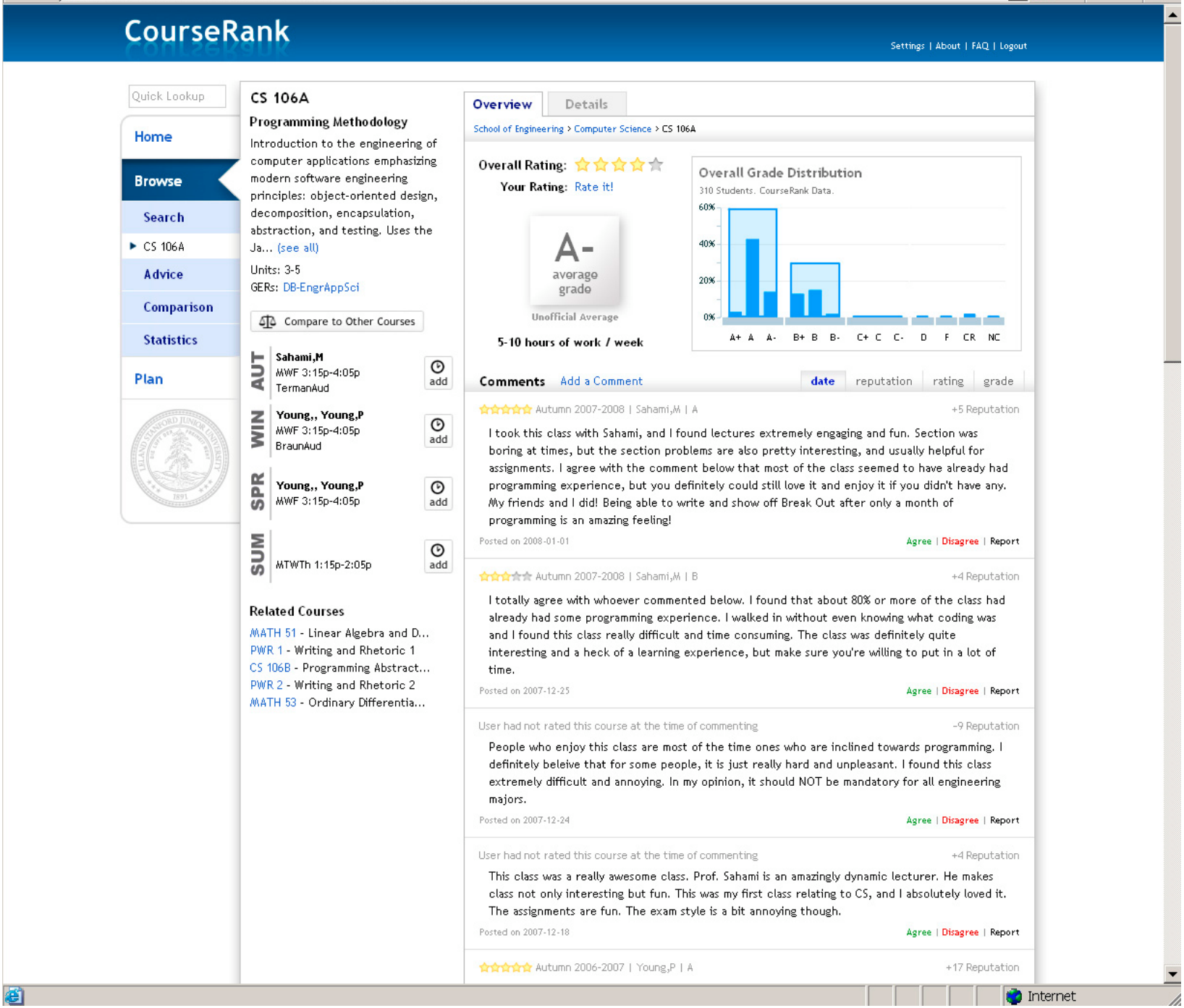


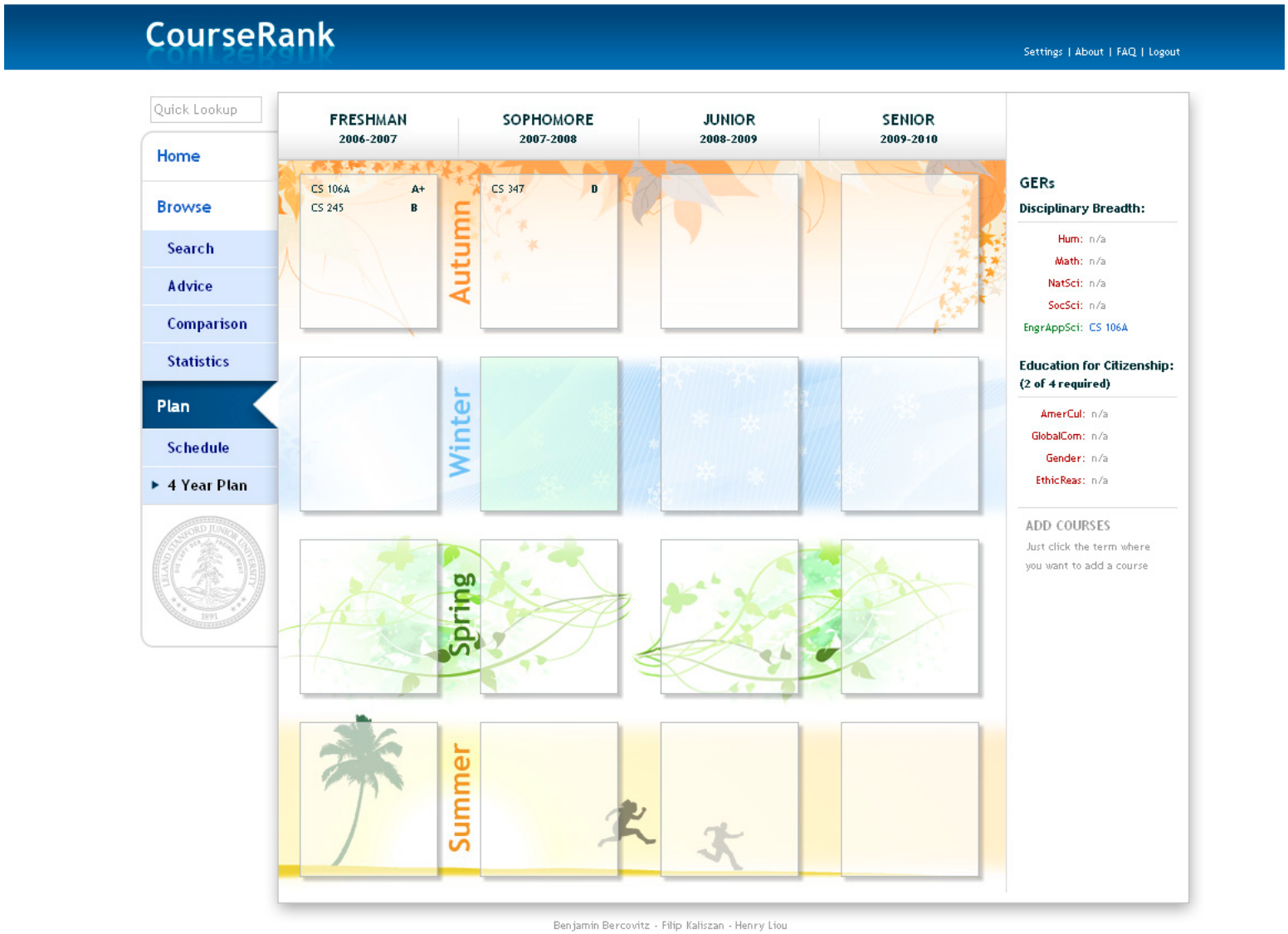
Stanford InfoLab Activities

- CourseRank
- PhotoSpread
- Tagging Systems
- Combating Spam
- Entity Resolution
- Sociologist's Workbench
- Privacy
- Uncertain Data
- and much more!

CourseRank


CourseRank - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Search Favorites
Address http://courserank.stanford.edu/CourseRank/
Go Links
CourseRank
About | FAQ | Login
Quick Course Lookup
welcome to the new
CourseRank
December 25, 2007
Have a nice break!
Autumn 2007 course ratings are now enabled
December 12, 2007
Rate from course page
Improved search & advice
TAs can now be students
December 9, 2007
2,000+ students
10,000+ ratings
Welcome
Browse Courses
Rate & Comment
Schedule & Plan
Track Requirements
Login
CourseRank: The Unofficial Course Guide
Stanford InfoLab
Now officially supported by Stanford University.
Over the past quarter, CourseRank joined forces with Stanford's InfoLab Group. Led by professor Hector Garcia-Molina and with the aid and cooperation of Stanford's administration / ASSU, a new and improved CourseRank is finally here.
Benjamin Bercovitz - Filip Kaliszan - Henry Liou


CourseRank
Settings | About | FAQ | Logout
Quick Lookup
Home
Browse
Search
CS 106A
Advice
Comparison
Statistics
Plan
CS 106A
Programming Methodology
Introduction to the engineering of computer applications emphasizing modern software engineering principles: object-oriented design, decomposition, encapsulation, abstraction, and testing. Uses the Ja... (see all)
Units: 3-5
GERs: DB-EngrAppSci
Compare to Other Courses
AUT Sahami,M MWF 3:15p-4:05p TermanAud add
WIN Young,, Young,P MWF 3:15p-4:05p BraunAud add
SPR Young,, Young,P MWF 3:15p-4:05p add
SUM MTWTh 1:15p-2:05p add
Related Courses
MATH 51 - Linear Algebra and D...
PWR 1 - Writing and Rhetoric 1
CS 106B - Programming Abstract...
PWR 2 - Writing and Rhetoric 2
MATH 53 - Ordinary Differentia...
Overview
Details
School of Engineering > Computer Science > CS 106A
Overall Rating:
Your Rating: Rate it!
A-
average grade
Unofficial Average
5-10 hours of work / week
Overall Grade Distribution
310 Students. CourseRank Data.
60%
40%
20%
0%
A+ A A- B+ B B- C+ C C- D F CR NC
Comments
Add a Comment
date
reputation
rating
grade
Autumn 2007-2008 | Sahami,M | A
+5 Reputation
I took this class with Sahami, and I found lectures extremely engaging and fun. Section was boring at times, but the section problems are also pretty interesting, and usually helpful for assignments. I agree with the comment below that most of the class seemed to have already had programming experience, but you definitely could still love it and enjoy it if you didn't have any. My friends and I did! Being able to write and show off Break Out after only a month of programming is an amazing feeling!
Posted on 2008-01-01
Agree | Disagree | Report
Autumn 2007-2008 | Sahami,M | B
+4 Reputation
I totally agree with whoever commented below. I found that about 80% or more of the class had already had some programming experience. I walked in without even knowing what coding was and I found this class really difficult and time consuming. The class was definitely quite interesting and a heck of a learning experience, but make sure you're willing to put in a lot of time.
Posted on 2007-12-25
Agree | Disagree | Report
User had not rated this course at the time of commenting
-9 Reputation
People who enjoy this class are most of the time ones who are inclined towards programming. I definitely beleive that for some people, it is just really hard and unpleasant. I found this class extremely difficult and annoying. In my opinion, it should NOT be mandatory for all engineering majors.
Posted on 2007-12-24
Agree | Disagree | Report
User had not rated this course at the time of commenting
+4 Reputation
This class was a really awesome class. Prof. Sahami is an amazingly dynamic lecturer. He makes class not only interesting but fun. This was my first class relating to CS, and I absolutely loved it. The assignments are fun. The exam style is a bit annoying though.
Posted on 2007-12-18
Agree | Disagree | Report
Autumn 2006-2007 | Young,P | A
+17 Reputation
Internet


CourseRank
Settings | About | FAQ | Logout
Quick Lookup
Home
Browse
Search
Advice
Comparison
Statistics
Plan
Schedule
4 Year Plan
FRESHMAN
2006-2007
SOPHOMORE
2007-2008
JUNIOR
2008-2009
SENIOR
2009-2010
CS 106A A+
CS 245 B
CS 347 D
Autumn
Winter
Spring
Summer
GERs
Disciplinary Breadth:
Hum: n/a
Math: n/a
NatSci: n/a
SocSci: n/a
EngrAppSci: CS 106A
Education for Citizenship:
(2 of 4 required)
AmerCul: n/a
GlobalCom: n/a
Gender: n/a
EthicReas: n/a
ADD COURSES
Just click the term where you want to add a course
Benjamin Bercovitz - Filip Kaliszan - Henry Liou

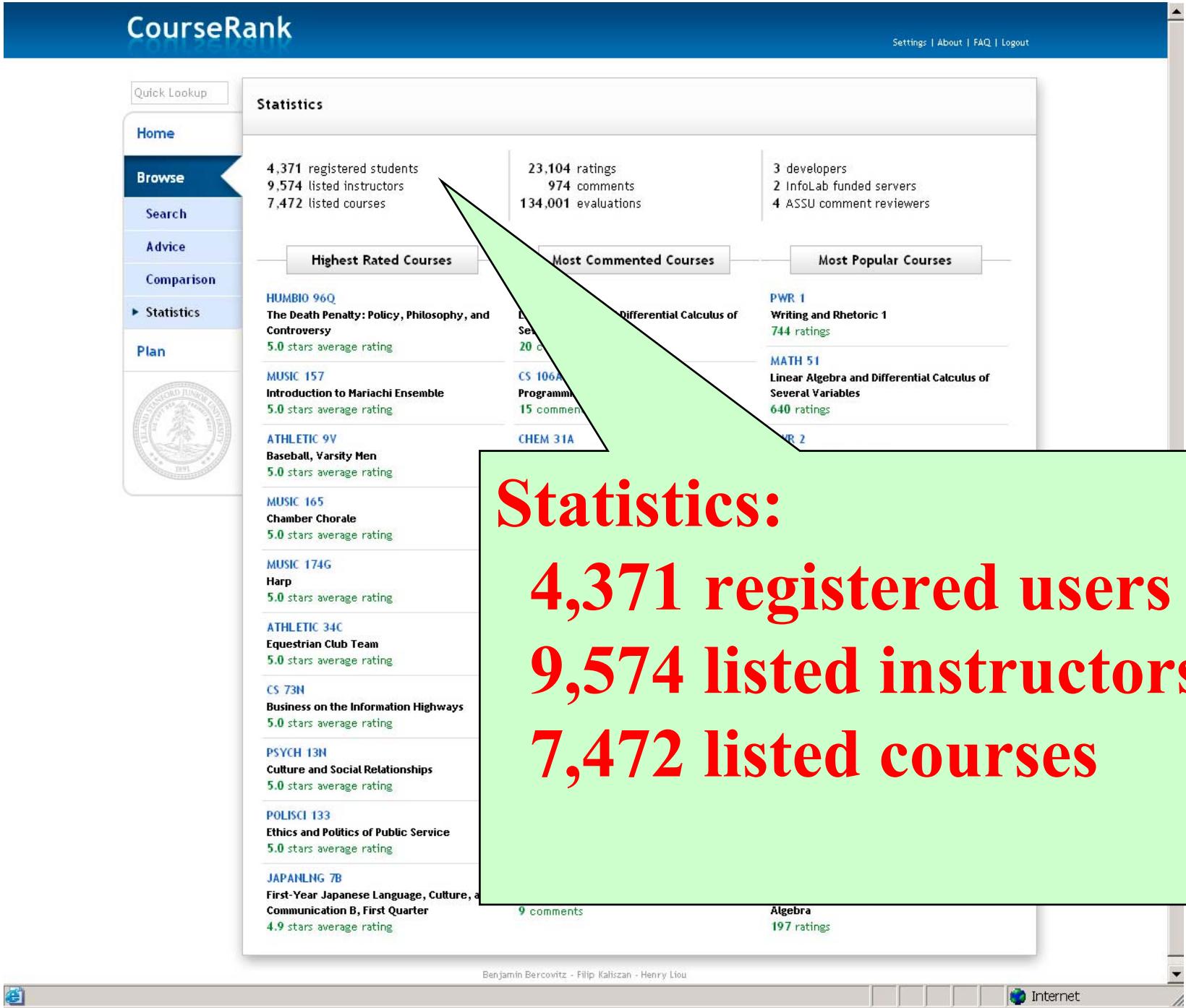
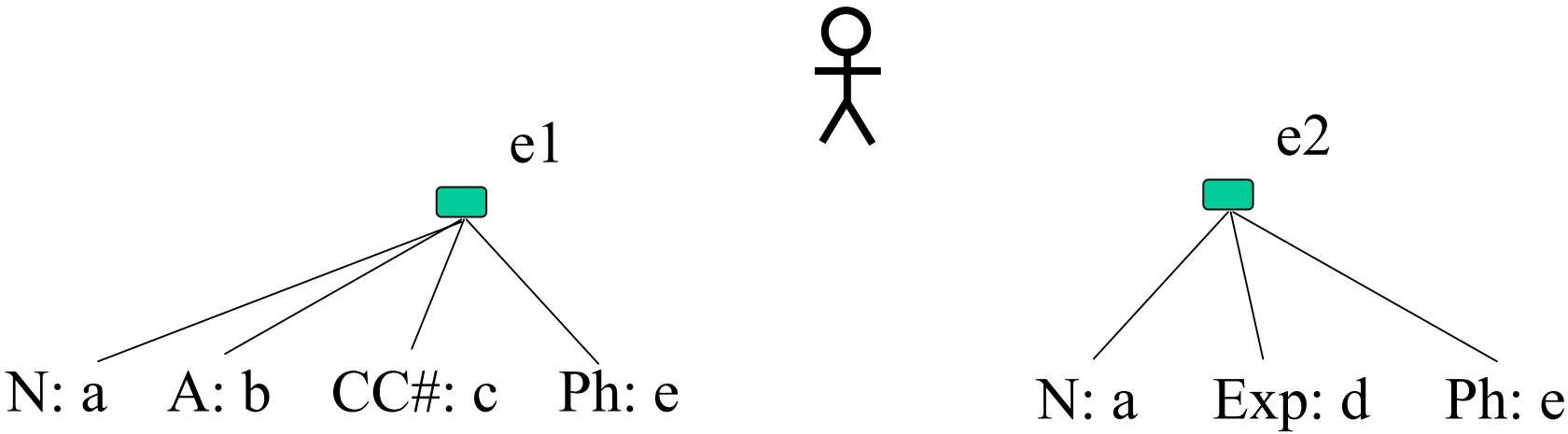

CourseRank
Settings | About | FAQ | Logout
Quick Lookup
Home
Browse
Search
Advice
Comparison
Statistics
Plan
Statistics
4,371 registered students
9,574 listed instructors
7,472 listed courses
23,104 ratings
974 comments
134,001 evaluations
3 developers
2 InfoLab funded servers
4 ASSU comment reviewers
Highest Rated Courses
Most Commented Courses
Most Popular Courses
HUMBIO 96Q
The Death Penalty: Policy, Philosophy, and Controversy
5.0 stars average rating
MUSIC 157
Introduction to Mariachi Ensemble
5.0 stars average rating
ATHLETIC 9V
Baseball, Varsity Men
5.0 stars average rating
MUSIC 165
Chamber Chorale
5.0 stars average rating
MUSIC 174G
Harp
5.0 stars average rating
ATHLETIC 34C
Equestrian Club Team
5.0 stars average rating
CS 73N
Business on the Information Highways
5.0 stars average rating
PSYCH 13N
Culture and Social Relationships
5.0 stars average rating
POLISCI 133
Ethics and Politics of Public Service
5.0 stars average rating
JAPANLNG 7B
First-Year Japanese Language, Culture, a
Communication B, First Quarter
4.9 stars average rating
CHEM 31A
9 comments
PWR 1
Writing and Rhetoric 1
744 ratings
MATH 51
Linear Algebra and Differential Calculus of Several Variables
640 ratings
Algebra
197 ratings
Statistics:
4,371 registered users
9,574 listed instructors
7,472 listed courses
Benjamin Bercovitz - Filip Kaliszan - Henry Liou
Internet

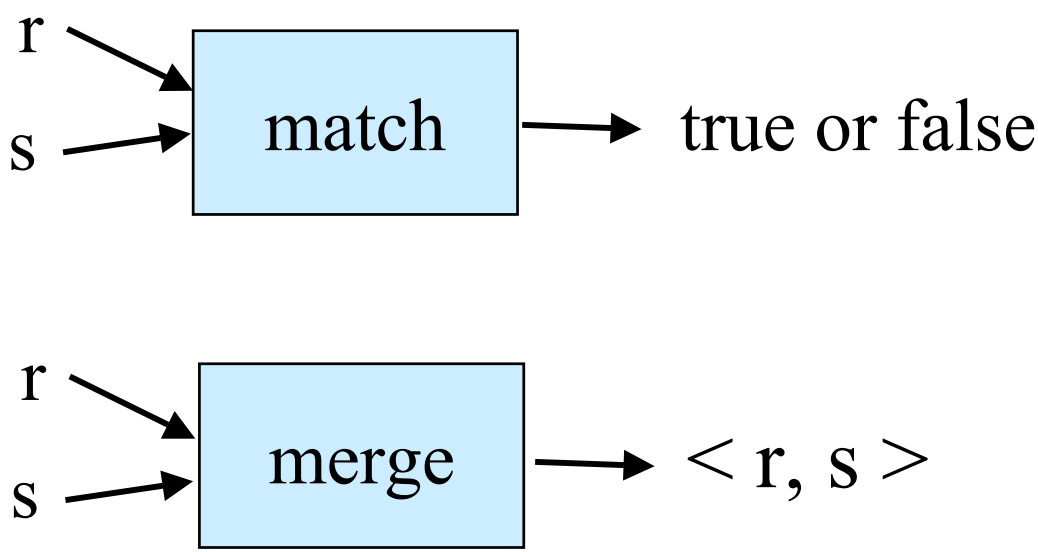
Photo SpreadSheet

PhotoSpread Interface

Entity Resolution



Entity Resolution at Stanford



Stanford
Lab