Information in the Social Web Era*

James Caverlee

caverlee@cse.tamu.edu
http://faculty.cse.tamu.edu/caverlee
HRBB 403
3rd Century BC

- Library of Alexandria
- 500,000 volumes
- catalogs and classifications
- “controlled vocabularies”

Based on Bruce Croft and Ned Fielden
Brief History of “Information”

48 BC
Brief History of “Information”

- First concordance of the Bible

- invention of the inverted list data structure

“River”:


Based on Bruce Croft and Ned Fielden
Brief History of “Information”

- University of Oxford Library 1600
  - all books printed in England
- Johnson’s Dictionary 1755
  - Set standard for dictionaries
  - Included common language
  - Helped standardize spelling
- Library of Congress 1800
- Webster’s Dictionary 1828
  - significantly larger than previous dictionaries
  - standardized American spelling
- Roget’s Thesaurus 1852
- Dewey Decimal classification 1876
- Carnegie Public Libraries 1880s
  - 1681 built (first public library 1850)

Based on Bruce Croft and Ned Fielden
Brief History of “Information”

- Vannevar Bush’s Memex 1945
- Personal information store based on associations

Based on Bruce Croft and Ned Fielden
Seriously ... my first laptop(?)

$3,000
<table>
<thead>
<tr>
<th>#.</th>
<th>Level: Name</th>
<th>Class</th>
<th>AC</th>
<th>Hits</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>(1+0)JOHN</td>
<td>Peasant</td>
<td>9</td>
<td>11/11</td>
</tr>
</tbody>
</table>

```
Option »»
```

Dungeon Options:
D, W, A, X - Move
Order, Spell,
Touch, Pty, Man,
Bury, Use, Quit,
G - Give (Xfer)
No Internet ... we had Sneakernet
Okay ... so the world has changed a lot
Not just devices ... also how they are connected: The Web is Ubiquitous
Millions of users
Anyone (almost) can use
Billions of Web pages, links, transactions, ...
→ the largest data source in the history of the world

How do we make sense of it all?
How to scale?
What algorithms, architectures?
Modeling and Analyzing the Web:

Content-Based Approaches

- Applications
  - Text-based Web retrieval
  - Web page categorization (e.g., News, Sports, Computers)
  - Clustering
  - …
Modeling and Analyzing the Web: Link-Based Approaches

- PageRank, HITS, ...
- Applications
  - Web page authority
  - Community detection
  - Spam detection
  - ...

Diagram showing a network model with nodes and links.
Explosion of the Social Web
Modeling and Analyzing the Web: User-Based Approaches

- Content and link approaches are **intrinsic** to the Web
- User-based approaches capture **external** (social, feedback, interaction) perspective on the Web

Examples:
- **Clicks**
- Comments
- Ratings
- Tweets
- Tags
- ...

query = “tamu”

<click>
User-Based Web Perspective: Comments

Amazon
Customer Reviews
5 star: (14)
4 star: (1)
3 star: (0)
2 star: (0)
1 star: (0)
Average Customer Review ★★★★★ (15 customer reviews)

NYTimes
READERS' COMMENTS
White House Scraps Bush's Approach to Missile Shield
By PETER BAKER
President Obama on Thursday announced a reconfigured system that won't be based in Poland or the Czech Republic, and will be aimed at intercepting Iranian missiles.
Comments are no longer being accepted.

Digg
Barack Obama wins the 2008 Presidential Election
36516 diggs
13311 Comments

Facebook
September 23 at 10:06am · Comment · Like · Share
Baiyu Chen, Enzo Wu and 2 others like this.
User-Based Web Perspective: Social Bookmarks

Large-Scale User-Driven Annotation of the Web
User-Based Web Perspective: Tweets

The new iPhone is great!

I'm visiting http://www.tamu.edu today
• **Our Research:**
  • Web-scale information management
  • Distributed data-intensive systems
  • Social computing.
  • Enable **efficient** and **trustworthy** information sharing and knowledge discovery over dynamic, heterogenous, and massive-scale **networked information systems**.
Two Major Research Thrusts

1. Spam-Resilient Web-Scale Computing
   - Predicting Web Spam with HTTP Session Information
     - Credibility-Based Link Analysis
   - Social Honeypots: Defending Online Communities from Spammers

2. Community-Based Mining and Information Management
   - SocialTrust: Enabling Trustworthy Information Access in Social Information Systems
   - Social Web Explorer: Analysis and Exploration of Socially-Generated Metadata
   - Learning Community-Preference of User-Contributed Comments
One Sample Project
Social Text Mining

Unigram language model

\[ \{p(t)\}_{t \in V} \text{ s.t. } \sum_{t \in V} p(t) = 1 \]

wrt to user-declared classes (e.g., male vs. female)
Which terms are most distinguishing for each class?

\[ MI(t, c) = p(t|c)p(c) \log \frac{p(t|c)}{p(t)} \]

\[ p(t|c) = \frac{\text{count}(c, t)}{\text{count}(c)} \quad \text{and} \quad p(c) = \frac{\text{count}(c)}{n} \]
By location?

camping, pdx, hiking, northwest, pixies, snowboarding, coast, rafting, floater, rad, wine, vegan.

catholic, yankees, nyc, uconn, hispanic, bronx, boston, sox, nas, italian, goodfellas, sneakers.
Hmm ...
User-Based Web Perspective: Tweets

The new iPhone is great!

I'm visiting http://www.tamu.edu today
Okay ... last bit of advice

- Go to talks
  - In the department
  - Across campus
- Own your education
  - Online opportunities
- Write, write, write
The Python Tutorial

Release: 2.6
Date: March 13, 2010

Python is an easy to learn, powerful programming language. It has efficient high-level data structures and a simple but effective approach to object-oriented programming. Python’s elegant syntax and dynamic typing, together with its interpreted nature, make it an ideal language for scripting and rapid application development in many areas on most platforms.

The Python interpreter and the extensive standard library are freely available in source or binary form for all major platforms from the Python Web site, http://www.python.org/, and may be freely distributed. The same site also contains distributions of and pointers to many free third party Python modules, programs and tools, and additional documentation.

The Python interpreter is easily extended with new functions and data types implemented in C or C++ (or other languages callable from C). Python is also suitable as an extension language for customizable applications.

This tutorial introduces the reader informally to the basic concepts and features of the Python language and system. It helps to have a Python interpreter handy for hands-on experience, but all examples are self-contained, so the tutorial can be read off-line as well.

For a description of standard objects and modules, see The Python Standard Library. The Python Language Reference gives a more formal definition of the language. To write extensions in C or C++, read Extending and Embedding the Python Interpreter and Python/C API Reference Manual. There are also several books covering Python in depth.

This tutorial does not attempt to be comprehensive and cover every single feature, or even every commonly used feature. Instead, it introduces many of Python’s most noteworthy features, and will give you a good idea of the

iPhone Application Development (Winter 2010)

Tools and APIs required to build applications for the iPhone platform using the iPhone SDK. User interface designs for mobile devices and unique user interactions using multitouch technologies. Object-oriented design using model-view controller pattern, memory management, Objective-C programming language, iPhone APIs and tools including Xcode, Interface Builder and Instruments on Mac OS X. Other topics include core animation,Bonjour networking, mobile device power management and performance considerations.

Prerequisites: C programming language experience at the level of 1568 or X. Recommended: UNIX, object-oriented programming, graphical tools.

Offered by Stanford's School of Engineering, the course will last ten weeks and include both the lecture videos and PDF documents. A new lecture will be posted a week after each class meeting. Subscribe to this course, and automatically receive new lectures as they become available.

Released with a Creative Commons BY-NC-ND License.
Google App Engine

Run your web apps on Google's infrastructure. Easy to build, easy to maintain, easy to scale.

Java™ Language Support
App Engine recently unveiled its second language: Java. This release includes our Java runtime, integration with Google Web Toolkit, and a Google Plugin for Eclipse, giving you an end-to-end Java solution for AJAX web applications. The Java runtime is now available for anyone to use, so please give it a try and send us your feedback.

- Get the full scoop in our blog post.
- Click over to YouTube to watch our Campfire One announcements.

Getting Started
1. Sign up for an App Engine account.
2. Download the App Engine SDK.
3. Read the Getting Started Guide.
4. Check out sample apps in the App Gallery.

Watch and Learn
Developing and deploying on Google App Engine. Watch Now

links from AppEngine.reddit.com

Pre-release 1.3.2 SDKs available
mobsoc.org

Mobile Applications • Social Media
Where social goes mobile
and where mobile gets social.
Where start-ups get academic
and where academics get started.
Where relationships get digital
and where profiles get relational.
Where clue gets personal,
and where personalities get clued.
Where communities get dissected
but never bisected.
It's where you share with us,
and where we share with you.

mobsoc is a place to talk about mobile computing,
social media, and/or the convergence of both.
Designers, developers, dreamers and PhDs are all
welcome.

mobsoc is a monthly meeting open to the public.
Attendees come from academia, the local startup
sxsw.com

TOMORROW HAPPENS HERE.

dena h boyd gives the Opening Remarks at the 2010 SXSW Interactive Festival.

What can we say? You came, and you brought it. We want to thank each and every one of you for helping to make SXSW® Interactive 2010 one of our best years ever. The panels, the parties, the 13th Annual Web Awards, the ScreenBurn at SXSW® Arcade, the Film and Interactive Trade Show and Exhibition, and Microsoft...
If you’re interested in these areas:
→ Good programming skills
→ Math: Probability, statistics, linear algebra
→ Machine learning, Info Retrieval, Databases

James Caverlee

caverlee@cse.tamu.edu
http://faculty.cse.tamu.edu/caverlee
HRBB 403

I am looking for talented and motivated undergraduates: email me if you’re interested!