The Era of Social Computing

Irwin King

Department of Computer Science and Engineering
The Chinese University of Hong Kong

king@cse.cuhk.edu.hk
http://www.cse.cuhk.edu.hk/~king

©2009 Irwin King. All rights reserved.
Sand from Centuries Past Send Future Voices Fast
Interdependence is and ought to be as much the ideal of man as self-sufficiency.

Man is a social being.

Mahatma Gandhi
A Brief History of the World
A Brief History of the World

1750
Industrial Revolution
ENIAC

1945
Information Age

1969
The MITS Altair

1975
Apple II

1981
Time Magazine Person of the Year

1983
Birth of WWW

1984
Birth of XML

1989
Birth of Web 2.0

1996
Attention Age

2004

2006

Billionaires’ Shuffle

Facebook in 2004.02

at **23** and **$1.5 billion** later...

**2008**

[Images of marked individuals]

Facebook’s Global Audience

Global Audience: 283,443,180

Data for 09/25/2009

Total Users | % Online Population
---|---

United States
Country Audience: 86,406,460
Percent of Global Audience: 30.48%

United States Male / Female

United States Age Distribution

Not Pictured: Hong Kong, Maldives, Palestine, Singapore, Taiwan
Facebook’s Growth Table

<table>
<thead>
<tr>
<th>10 Largest Countries</th>
<th>10 Fastest Growing Over Past Week</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. United States</td>
<td>1. China</td>
</tr>
<tr>
<td>2. United Kingdom</td>
<td>2. Taiwan</td>
</tr>
<tr>
<td>3. Turkey</td>
<td>3. Vietnam</td>
</tr>
<tr>
<td>4. Canada</td>
<td>4. Philippines</td>
</tr>
<tr>
<td>5. France</td>
<td>5. Iraq</td>
</tr>
<tr>
<td>6. Italy</td>
<td>6. Romania</td>
</tr>
<tr>
<td>7. Indonesia</td>
<td>7. Sweden</td>
</tr>
<tr>
<td>8. Australia</td>
<td>8. Ireland</td>
</tr>
<tr>
<td>10. Argentina</td>
<td>10. Qatar</td>
</tr>
</tbody>
</table>

General Growth: More than 300 million active users
50% of our active users log on to Facebook in any given day
The fastest growing demographic is those 35 years old and older
## Global Internet Traffic

<table>
<thead>
<tr>
<th>Alexa as of May 2009</th>
<th>China</th>
<th>USA</th>
<th>Japan</th>
<th>India</th>
<th>Brazil</th>
<th>Global</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Baidu</td>
<td>Google</td>
<td>Yahoo.jp</td>
<td>Google.in</td>
<td>Google</td>
<td>Google</td>
</tr>
<tr>
<td>2</td>
<td><strong>QQ</strong></td>
<td>Yahoo</td>
<td><strong>FC2</strong></td>
<td>Google</td>
<td><strong>Orkut.br</strong></td>
<td>Yahoo</td>
</tr>
<tr>
<td>3</td>
<td>Sina</td>
<td>Facebook</td>
<td>Google.jp</td>
<td>Yahoo</td>
<td>Windows Live</td>
<td>YouTube</td>
</tr>
<tr>
<td>4</td>
<td>Google.cn</td>
<td><strong>YouTube</strong></td>
<td><strong>YouTube</strong></td>
<td><strong>Orkut.in</strong></td>
<td>Universo Online</td>
<td><strong>Facebook</strong></td>
</tr>
<tr>
<td>5</td>
<td>Taobao</td>
<td><strong>Myspace</strong></td>
<td>Rakuten</td>
<td><strong>YouTube</strong></td>
<td><strong>YouTube</strong></td>
<td>Windows Live</td>
</tr>
<tr>
<td>6</td>
<td>163</td>
<td>MSN</td>
<td>Livedoor</td>
<td><strong>Blogger</strong></td>
<td>Globo</td>
<td>MSN</td>
</tr>
<tr>
<td>7</td>
<td>Google</td>
<td>Windows Live</td>
<td><strong>Ameblo.jp</strong></td>
<td>Rediff</td>
<td>MSN</td>
<td><strong>Wikipedia</strong></td>
</tr>
<tr>
<td>8</td>
<td>Sohu</td>
<td><strong>Wikipedia</strong></td>
<td><strong>mixi</strong></td>
<td><strong>Facebook</strong></td>
<td>Google</td>
<td><strong>Blogger</strong></td>
</tr>
<tr>
<td>9</td>
<td>Youku</td>
<td>Craigslist</td>
<td><strong>Wikipedia</strong></td>
<td><strong>Wikipedia</strong></td>
<td>Yahoo</td>
<td>Baidu</td>
</tr>
<tr>
<td>10</td>
<td>Yahoo</td>
<td>EBay</td>
<td>Google</td>
<td>Windows Live</td>
<td>Terra</td>
<td><strong>Myspace</strong></td>
</tr>
</tbody>
</table>
Twitter in Spotlight

The Lede
The New York Times News Blog
June 2, 2009, 7:05 PM
China's Great Firewall Blocks Twitter
By ROBERT MACKEY

Recent Posts
June 18
(38 comments)
Latest Updates on Iran's Disputed Election
To supplement reporting from New York Times correspondents inside Iran on Thursday, The Lede will continue to track the aftermath of Iran's disputed presidential election online.

June 17
(129 comments)
Wednesday: Latest Updates on Iran's Disputed Election
On Wednesday, The Lede will continue to track the aftermath of Iran's disputed presidential election online, to supplement reporting from New York Times correspondents inside Iran.

June 16
(198 comments)
Tuesday: Latest Updates on Iran's Disputed Election
To supplement reporting from New York Times correspondents inside Iran, The Lede
Web 2.0

- Web as a medium vs. **Web as a platform**
- Read-Only Web vs. **Read-and-Write Web**
- Static vs. **Dynamic**
- Restrictive vs. **Freedom & Empowerment**
- Technology-centric vs. **User-centric**
- Limited vs. **Rich User Experience**
- Individualistic vs. **Group/Collective Behavior**
- Consumer vs. **Producer**
- Transactional vs. **Relational**
- Top-down vs. **Bottom-up**
- People-to-Machine vs. **People-to-People**
- Search & browse vs. **Publish & Subscribe**
- Closed application vs. **Service-oriented Services**
- Functionality vs. **Utility**
- Data vs. **Value**
Social Networks

Society:
Nodes: individuals
Links: social relationship (family/work/friendship/etc.)

S. Milgram and John Guare: *Six Degree of Separation.*
Social networks: Many *individuals* with diverse *social interactions* between them.
Milgram’s Experiment
Social Networks

- The Earth is developing an electronic nervous system, a network with diverse **nodes** and **links**.

Communication networks: many non-identical components with diverse connections between them.
The Flow of Information
Examples

[Kleinberg 1999]
Organizational Chart
Social Network Chart

Authority vs. Importance

<table>
<thead>
<tr>
<th>Rank</th>
<th>Name</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>Nancy</td>
<td>(secretary)</td>
</tr>
<tr>
<td>0.66</td>
<td>Donna</td>
<td>(supervisor)</td>
</tr>
<tr>
<td>0.57</td>
<td>Manuel</td>
<td>(manager)</td>
</tr>
<tr>
<td>0.19</td>
<td>Stuart</td>
<td>(supervisor)</td>
</tr>
<tr>
<td>0.17</td>
<td>Charles</td>
<td>(supervisor)</td>
</tr>
<tr>
<td>0.08</td>
<td>Kathy</td>
<td>(secretary)</td>
</tr>
<tr>
<td>0.08</td>
<td>Tanya</td>
<td>(secretary)</td>
</tr>
<tr>
<td>0.02</td>
<td>Fred</td>
<td>(auditor)</td>
</tr>
<tr>
<td>0.00</td>
<td>Sharon</td>
<td>(auditor)</td>
</tr>
<tr>
<td>0.00</td>
<td>Bob</td>
<td>(auditor)</td>
</tr>
<tr>
<td>0.00</td>
<td>Carol</td>
<td>(auditor)</td>
</tr>
<tr>
<td>0.00</td>
<td>Harold</td>
<td>(auditor)</td>
</tr>
<tr>
<td>0.00</td>
<td>Wynn</td>
<td>(auditor)</td>
</tr>
<tr>
<td>0.00</td>
<td>Susan</td>
<td>(secretary)</td>
</tr>
</tbody>
</table>

Social Networking Sites

- Example of Social Networking Sites: Facebook, MySpace, Blogger, QQ, etc.
Social Search

- Social Search Engine
- Leveraging your social networks for searching
Social Media

- YouTube
  - Promoted Videos: Think Again Awards, Think Again Awards, Think Again Awards
  - Featured Videos: David Sedaris delivers a pizza, Erbert and Gerbert’s Candle Cannon
- Flickr
  - Featured Videos: iLike
- iLike
  - Social networking platform

Social News/Mash Up

On-line Games and Virtual Communities
Social Bookmarking

Social Entertainment

[Image of a website interface promoting various auctions, including a 300 Bids Voucher, MySims Agents (Nintendo DS), Samsung UN46B6000 46-Inch 1080p LED HDTV, Wii | Nintendo Console + Wii Sports, and Apple MacBook Pro MB991LL/A 13.3-Inch Laptop.]
Social Recommendations

Genius Recommendations for Apps
There are tens of thousands of apps in the App Store, with more added every day. A new feature of iPod touch makes finding cool new apps even easier. It's Genius for apps, and it works just like Genius for your music. Tap the Genius icon and get recommendations for apps that you might like based on apps you and others have downloaded.

Genius Playlists
Say you're listening to a song you really like and want to hear other tracks that go great with it. The Genius feature finds other songs on your iPod touch that sound great with the one you were listening to and makes a Genius playlist for you. Listen to the playlist right away, save it for later, or even refresh it and give it another go. Count on Genius to create a mix you wouldn't have thought of yourself.

Genius Mixes
Now the Genius feature is even more powerful. Introducing Genius Mixes. All you do is sync iPod touch to iTunes, and Genius automatically searches your library to find songs that sound great together. Then it creates multiple mixes you'll love. These mixes are like channels programmed entirely with your music.
The notion of social informatics relates to the interaction between society and ICT (information-communication technologies). In its broadest sense it covers:

1. the social consequences of ICT at micro (e.g. social aspects of ICT applications at personal and organisational level) as well as at macro level (e.g. information society studies);
2. the application of ICT in the area of social sciences and social/public sector;
3. the use of ICT as a tool for studying social phenomena (within social science methodology).

Graphical presentation is here>>

News

07.12.09  Information Society Free Virtual Library
02.12.09  Job offer: Professor in Social Informatics
01.12.09  Call for papers to "New technologies and data collection in social sciences"
09.11.09  Call for Papers "IASSIST 2010"
27.10.09  Job offer: Associate Professor Position - Department of Social Informatics

Social Knowledge Sharing
Social Marketing

- Viral marketing
- Who are the brokers?
- Who can exert the most influence on buying/selling?
- How much should one advertise?
Social/Human Computation
Chinese CAPTCHA

Ling-Jyh Chen, Institute of Information Science, Academia Sinica, Taipei, Taiwan
Human Computation
Games With A Purpose

- **Matchin**
  - Image search by aesthetic value

- **Babble**
  - Translate foreign language into English

- **InTune**
  - Tags songs with description text

- **Squigl**
  - Image segmentation

- **Verbosity**
  - Database of common knowledge description
Web 2.0 Revolution

- **Glocalization** - think globally and act locally!
- **Weblication** - Web is the application!
- **Three C’s**
  - Connectivity
  - Collaboration
  - Communities
Topics in Social Computing

- Social Behavior Analysis and Modeling
- Social Media
- Social Network Theory and Models
- Link Analysis/Graph Mining/Large Graph Algorithms
- Recommender Systems/Collaborative Filtering
- QA/Sentiment Analysis/Opinion Mining

- Human Computation/Crowdsourcing
- Risk, Trust, Security, and Privacy
- Monetization of Social Computing
- Software Tools and Applications
- and many, many more...
Social Computing

Online communities
- Blogs
- Wikis
- Social networks
- Collaborative bookmarking
- Social tagging
- Podcasts

Business and public sector
- Recommendation
- Forecasting
- Reputation
- Feedback
- Decision analysis
- E-government

Interactive entertainment
- Edutainment
- Training
- Gaming
- Storytelling

Technological infrastructure
- Web technology
- Database technology
- Multimedia technology
- Wireless technology
- Agent technology
- Software engineering

Theoretical underpinnings
- Social psychology
- Communication and human-computer interaction theories
- Social network analysis
- Anthropology
- Organization theory
- Sociology
- Computing theory

[Wang et al. 2007]
Definition of Social Computing

• Any Computer-mediated communication and interaction

• In the weaker sense: supporting any sort of social behavior
  • blogs, email, instant messaging, wiki, social network services, social bookmarking

• In the stronger sense: supporting “computations” that are carried out by a group of people
  • collaborative filtering, online auctions, prediction markets, reputation systems, tagging, verification games
Motivation

1. Difficult for users to express information needed
2. Word mismatch in information retrieval
Motivation

1. Accurate to express information needed
2. Easy to inform information
Challenges

• Queries contain **ambiguous and new terms**
  
  • apple: “apple computer” or “apple pie”?
  
  • NDCG:?  

• Users tend to submit **short queries** consisting of only one or two words  
  
  • almost 20% one-word queries
  
  • almost 30% two-word queries

• Users may have **little or even no knowledge** about the topic they are searching for!
Query Suggestion Using Clickthrough Data

- Query logs recorded by search engines

\[ \langle u, q, l, r, t \rangle \]

<table>
<thead>
<tr>
<th>ID</th>
<th>Query</th>
<th>URL</th>
<th>Rank</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>358</td>
<td>facebook</td>
<td><a href="http://www.facebook.com">http://www.facebook.com</a></td>
<td>1</td>
<td>2008-01-01 07:17:12</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

- Users’ relevance feedback to indicate desired/preferred/target results
Joint Bipartite Graph

$B_{uq} = (V_{uq}, E_{uq})$

$V_{uq} = U \cup Q$

$U = \{u_1, u_2, ..., u_m\}$

$Q = \{q_1, q_2, ..., q_n\}$

$E_{uq} = \{(u_i, q_j) | \text{there is an edge from } u_i \text{ to } q_j\}$

is the set of all edges.

The edge $(u_i, q_j)$ exists in this bipartite graph if and only if a user $u_i$ issued a query $q_j$.

$B_{ql} = (V_{ql}, E_{ql})$

$V_{ql} = Q \cup L$

$Q = \{q_1, q_2, ..., q_n\}$

$L = \{l_1, l_2, ..., l_p\}$

$E_{ql} = \{(q_i, l_j) | \text{there is an edge from } q_i \text{ to } l_j\}$

is the set of all edges.

The edge $(q_j, l_k)$ exists if and only if a user $u_i$ clicked a URL $l_k$ after issuing an query $q_j$. 
Key Points

- Two-level latent semantic analysis
  - Consider the use of a joint *user-query* and *query-URL bipartite graphs* for query suggestion
  - Use *matrix factorization* for learning query features in constructing the Query Similarity Graph
  - Use *heat diffusion* for similarity propagation for query suggestions
- Queries are issued by the users, and which URLs to click are also decided by the users.

- Two distinct users are similar if they issued similar queries.

- Two queries are similar if they are issued by similar users.
\(r_{ij}^*\)  Normalized weight, how many times \(u_i\) issued \(q_j\)

\(s_{jk}^*\)  Normalized weight, how many times \(q_j\) is linked to \(l_k\)

\(U_i\)  \(L\)-dimensional vector of user \(u_i\)

\(Q_j\)  \(L\)-dimensional vector of query \(q_j\)

\(L_k\)  \(L\)-dimensional vector of URL \(l_k\)

\[
\mathcal{H}(R, U, Q) = \min_{U, Q} \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} I_{ij}^R (r_{ij}^* - g(U_i^T Q_j))^2 + \frac{\alpha_u}{2} \|U\|_F^2 + \frac{\alpha_q}{2} \|Q\|_F^2
\]

\[
\mathcal{H}(S, Q, L) = \min_{Q, L} \frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{p} I_{jk}^S (s_{jk}^* - g(Q_j^T L_k))^2 + \frac{\alpha_q}{2} \|Q\|_F^2 + \frac{\alpha_l}{2} \|L\|_F^2
\]
A local minimum can be found by performing gradient descent in $U_i$, $Q_j$ and $L_k$.
Gradient Descent Equations

\[
\frac{\partial H}{\partial U_i} = \alpha_r \sum_{j=1}^{n} I_{ij}^R g'(U_i^T Q_j)(g(U_i^T Q_j) - r_{ij}^*) Q_j + \alpha_u U_i,
\]

\[
\frac{\partial H}{\partial Q_j} = \sum_{k=1}^{p} I_{jk}^S g'(Q_j^T L_k)(g(Q_j^T L_k) - s_{jk}^*) L_k
\]

\[+ \alpha_r \sum_{i=1}^{m} I_{ij}^R g'(U_i^T Q_j)(g(U_i^T Q_j) - r_{ij}^*) U_i + \alpha_q Q_j,
\]

\[
\frac{\partial H}{\partial L_k} = \sum_{j=1}^{n} I_{jk}^S g'(Q_j^T L_k)(g(Q_j^T L_k) - s_{jk}^*) Q_j + \alpha_l L_k,
\]

Only the \textbf{Q matrix}, the queries’ latent features, is being used to generate the \textbf{query similarity graph}!
Query Similarity Graph

- Similarities are calculated using queries’ latent features
- Only the top-$k$ similar neighbors (terms) are kept

$k = 4$
Similarity Propagation

• Based on the **Heat Diffusion Model**

• In the query graph, given the **heat sources** and the **initial heat values**, start the heat diffusion process and perform **$P$ steps**

• Return the **Top-$N$** queries in terms of highest heat values for query suggestions
Heat Diffusion Model

- Heat diffusion is a **physical phenomena**
- Heat flows from **high** temperature to **low** temperature in a **medium**
- **Heat kernel** is used to describe the amount of heat that one point receives from another point
- The way that heat diffuse varies when the **underlying geometry** varies

$$\rho C_P \frac{\partial T}{\partial t} = Q + \nabla \cdot (k \nabla T)$$

- $\rho$: Density
- $C_P$: Heat capacity and constant pressure
- $\frac{\partial T}{\partial t}$: Change in temperature over time
- $Q$: Heat added
- $k$: Thermal conductivity
- $\nabla T$: Temperature gradient
- $\nabla \cdot v$: Divergence
Heat Diffusion Process
Similarity Propagation Model

\[ \frac{f_i(t + \Delta t) - f_i(t)}{\Delta t} = \alpha \left( -\frac{\tau_i}{d_i} f_i(t) \sum_{k: (q_i, q_k) \in E} w_{ik} + \sum_{j: (q_j, q_i) \in E} \frac{w_{ji}}{d_j} f_j(t) \right) \] (1)

\[ f(1) = e^{\alpha H} f(0) \] (2)

\[ H_{ij} = \begin{cases} w_{ji}/d_j, & (q_j, q_i) \in E, \\ -\left(\tau_i/d_i\right) \sum_{k: (i, k) \in E} w_{ik}, & i = j, \\ 0, & \text{otherwise.} \end{cases} \] (3)

\[ f(1) = e^{\alpha R} f(0), \quad \mathbf{R} = \gamma H + (1 - \gamma) \mathbf{g} \mathbf{1}^T \] (4)

- \( \alpha \): Thermal conductivity
- \( d_i \): Heat value of node \( i \) at time \( t \)
- \( f_i(t) \): Heat value of node \( i \) at time \( t \)
- \( w_{ik} \): Weight between node \( i \) and node \( k \)
- \( f(0) \): Vector of the initial heat distribution
- \( f(1) \): Vector of the heat distribution at time 1
- \( \tau_i \): Equal to 1 if node \( i \) has outlinks, else equal to 0
- \( \gamma \): Random jump parameter, and set to 0.85
- \( \mathbf{g} \): Uniform stochastic distribution vector
Discrete Approximation

• Compute $e^{\alpha R}$ is time consuming

• We use the **discrete approximation** to substitute

$$f(1) = \left( I + \frac{\alpha}{P} R \right)^P f(0)$$

• For every heat source, only diffuse heat to its neighbors within $P$ steps

• In our experiments, $P = 3$ already generates fairly good results
Query Suggestion Procedure

- For a given query $q$

1. Select a set of $n$ queries, each of which contains at least one word in common with $q$, as heat sources

2. Calculate the initial heat values by

$$f_{\hat{q}_i}(0) = \frac{|\mathcal{W}(q) \cap \mathcal{W}(\hat{q}_i)|}{|\mathcal{W}(q) \cup \mathcal{W}(\hat{q}_i)|}$$

3. Use $f(1) = e^{\alpha R} f(0)$ to diffuse the heat in graph

4. Obtain the Top-N queries from $f(1)$
Physical Meaning of $\alpha$

- If set $\alpha$ to a large value
  - The results depend more on the query graph, and more semantically related to original queries, e.g., travel => lowest air fare

- If set $\alpha$ to a small value
  - The results depend more on the initial heat distributions, and more literally similar to original queries, e.g., travel => travel insurance
## Experimental Dataset

<table>
<thead>
<tr>
<th>Data Source</th>
<th>Clickthrough data from AOL search</th>
<th>After Pre-Processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Collection Period</td>
<td>March 2006 to May 2006 (3 months)</td>
<td></td>
</tr>
<tr>
<td>Lines of Logs</td>
<td>19,442,629</td>
<td></td>
</tr>
<tr>
<td>Unique user IDs</td>
<td>657,426</td>
<td>192,371</td>
</tr>
<tr>
<td>Unique queries</td>
<td>4,802,520</td>
<td>224,165</td>
</tr>
<tr>
<td>Unique URLs</td>
<td>1,606,326</td>
<td>343,302</td>
</tr>
<tr>
<td>Unique words</td>
<td></td>
<td>69,937</td>
</tr>
</tbody>
</table>
Pre-processing

• Computer set-up
  Intel Pentium D CPU, 3.0 Gz, Dual Core with 1G memory

• Keep **valid** words which contains only ‘a’, ‘b’,…, ‘z’ and spaces

• Remove those queries which appear less than **three** times
# Query Suggestions

Table 2: Examples of LSQS Query Suggestion Results ($k = 50$)

<table>
<thead>
<tr>
<th>Testing Queries</th>
<th>Suggestions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\alpha = 10$</td>
</tr>
<tr>
<td></td>
<td>Top 1</td>
</tr>
<tr>
<td>michael jordan</td>
<td>michael jordan shoes</td>
</tr>
<tr>
<td>travel</td>
<td>travel insurance</td>
</tr>
<tr>
<td>java</td>
<td>sun java</td>
</tr>
<tr>
<td>global services</td>
<td>ibm global services</td>
</tr>
<tr>
<td>walt disney land</td>
<td>world of disney</td>
</tr>
<tr>
<td>intel</td>
<td>intel vs amd</td>
</tr>
<tr>
<td>job hunt</td>
<td>jobs in maryland</td>
</tr>
<tr>
<td>photography</td>
<td>photography classes</td>
</tr>
<tr>
<td>internet explorer</td>
<td>ms internet explorer</td>
</tr>
<tr>
<td>fitness</td>
<td>fitness magazine</td>
</tr>
<tr>
<td>m schumacher</td>
<td>schumacher</td>
</tr>
<tr>
<td>solar system</td>
<td>solar system project</td>
</tr>
<tr>
<td>sunglasses</td>
<td>replica sunglasses</td>
</tr>
<tr>
<td>search engine</td>
<td>audio search engine</td>
</tr>
<tr>
<td>disease</td>
<td>grovers disease</td>
</tr>
<tr>
<td>pizzahut</td>
<td>pizza hut menu</td>
</tr>
<tr>
<td>health care</td>
<td>health care proxy</td>
</tr>
<tr>
<td>flower delivery</td>
<td>global flower delivery</td>
</tr>
<tr>
<td>wedding</td>
<td>wedding guide</td>
</tr>
<tr>
<td>astronomy</td>
<td>astronomy magazine</td>
</tr>
</tbody>
</table>
Emerging Issues

- **Theory** and models
- **Search, mining, and ranking** of existing information, e.g., **spatial** (relations) and **temporal** (time) domains
- Dealing with **partial and incomplete** information, e.g., collaborative filtering, ranking, tagging, etc.
- **Scalability** and **algorithmic** issues
- **Security, privacy, trust, and risk** issues
- **Monetization** of social interactions
- **Software platforms** and development **tools**
<table>
<thead>
<tr>
<th>Tool</th>
<th>Use now</th>
<th>Within five years</th>
<th>Don’t know/Not applicable</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blogs</td>
<td>44</td>
<td>32</td>
<td>24</td>
</tr>
<tr>
<td>Wikis</td>
<td>41</td>
<td>30</td>
<td>29</td>
</tr>
<tr>
<td>Mashups</td>
<td>10</td>
<td>25</td>
<td>66</td>
</tr>
<tr>
<td>Video podcasts</td>
<td>✔️</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Online courses</td>
<td>✔️</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social networks</td>
<td>✔️</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Text messaging/notifications</td>
<td>✔️</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collaboration software</td>
<td>✔️</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Document management</td>
<td>✔️</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RFID/sensor networks</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mobile broadband</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other, please specify</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Concluding Remarks

• The Era of Social Computing is here to stay!

• Relations are important!

• Discovering new paradigms by blending different social media and interactions

• Be concerned about computational techniques to search, rank, and mine data and information to achieve collective intelligence/wisdom
Acknowledgments

• Prof. Michael Lyu
• Mr. Patrick Lau
• Mr. Lam Cho Fung
• Mr. Simon Mok
• Mr. Ivan Yau
• Ms. Sara Fok
• Hongbo Deng (Ph.D.)
• Baichuan Li (M.Phil.)

• Zhenjiang Lin (Ph.D.)
• Hao Ma (Ph.D.)
• Mingzhe Mo (M.Phil.)
• Dingyan Wang (M.Phil.)
• Wei Wang (M.Phil.)
• Haiqin Yang (Ph.D.)
• Connie Yuen (Ph.D.)
• Xin Xin (Ph.D.)
• Chao Zhou (Ph.D.)
• Yi Zhu (Ph.D.)
On-Going Research

Machine Learning

• Heavy-Tailed Symmetric Stochastic Neighbor Embedding (NIPS’09)
• Adaptive Regularization for Transductive Support Vector Machine (NIPS’09)
• Direct Zero-norm Optimization for Feature Selection (ICDM’08)
• Semi-supervised Learning from General Unlabeled Data (ICDM’08)
• Learning with Consistency between Inductive Functions and Kernels (NIPS’08)
• An Extended Level Method for Efficient Multiple Kernel Learning (NIPS’08)
• Semi-supervised Text Categorization by Active Search (CIKM’08)
• Transductive Support Vector Machine (NIPS’07)
• Global and local learning (ICML’04, JMLR’04)
On-Going Research

Web Intelligence/Information Retrieval

• A Generalized Co-HITS Algorithm and Its Application to Bipartite Graphs (KDD’09)
• Entropy-biased Models for Query Representation on the Click Graph (SIGIR’09)
• Effective Latent Space Graph-based Re-ranking Model with Global Consistency (WSDM’09)
• Formal Models for Expert Finding on DBLP Bibliography Data (ICDM’08)
• Learning Latent Semantic Relations from Query Logs for Query Suggestion (CIKM’08)
• RATE: a Review of Reviewers in a Manuscript Review Process (WI’08)
• MatchSim: link-based web page similarity measurements (WI’07)
• Diffusion rank: Ranking web pages based on heat diffusion equations (SIGIR’07)
• Web text classification (WWW’07)
On-Going Research

Recommender Systems/Collaborative Filtering

• Learning to Recommend with Social Trust Ensemble (SIRIR’09)
• Semi-Nonnegative Matrix Factorization with Global Statistical Consistency in Collaborative Filtering (CIKM’09)
• Recommender system: accurate recommendation based on sparse matrix (SIGIR’07)
• SoRec: Social Recommendation Using Probabilistic Matrix Factorization (CIKM’08)

Human Computation

• A Survey of Human Computation Systems (SCA’09)
• Mathematical Modeling of Social Games (SIAG’09)
• An Analytical Study of Puzzle Selection Strategies for the ESP Game (WI’08)
• An Analytical Approach to Optimizing The Utility of ESP Games (WI’08)
Welcome to the workshop on Social Computing in Education (SCE2009). The workshop is held in conjunction with the SocialComp-09, Vancouver, Canada from August 29-31, 2009.

With the advent of Web 2.0 and related technologies, Social Computing has become a new paradigm in ways we communicate, learn, and educate. Social platforms such as wikis, blogs, twitters, forums, groups, podcasts, mashups, virtual worlds, and sites for social networking, recommender systems, social bookmarking, social news, knowledge sharing, etc. are generating novel ways we acquire, access, manipulate, process, retrieve, present, and visualize information in the teaching and learning space. The social media for education has become dynamic, ubiquitous, distributed, real-time, collaborative, bottom-up, many-to-many, value-based, and personalized. This workshop solicits contributions on using Social Computing and related technologies for education, the emerging applications of Web 2.0 as an educational platform, as well as privacy, risk, security, and policy issues associated in Social Computing for Education 2.0.
Since its inception, the Web has changed the landscape of human experiences on how we interact with one another and data through service infrastructures via various computing devices. This interweaving environment is now becoming ever more embedded into devices and systems that integrate seamlessly into how we live, both in our working or leisure time.

For this volume, King and Baeza-Yates selected some pioneering and cutting-edge research work that is pointing to the future of the Web. Based on the Workshop Track of the 17th International World Wide Web Conference (WWW2008) in Beijing, they selected the top contributions and asked the authors to resubmit their work with a minimum of one third of additional material from their original workshop manuscripts to be considered for this volume. After a second-round of reviews and selection, 16 contributions were finally accepted.

The work within this volume represents the tip of an iceberg of the many exciting advancements on the WWW. It covers topics like semantic web services, location-based and mobile applications, personalized and context-dependent user interfaces, social networks, and folksonomies. The presentations aim at researchers in academia and industry by showcasing latest research findings. Overall they deliver an excellent picture of the current state-of-the-art, and will also serve as the basis for ongoing research discussions and point to new directions.
In what ways do new technologies pose the greatest challenges and risks to colleges and universities? Select up to three.

(% of respondents)

Potential increase in student plagiarism

51
VeriGuide

- **Similarity text** detection system
- Developed at **CUHK**
- Promote and uphold academic **honesty, integrity, and quality**
- Support **English, Traditional** and **Simplified Chinese**
- Handle **.doc, .txt, .pdf, .html, etc. file formats**
- Generate detailed **originality report** including **readability**
VeriGuide Free Trial

http://www.cse.cuhk.edu.hk/~king

Q & A