Introduction to Social Computing

Irwin King

ATT Labs, Research
&
Department of Computer Science and Engineering
The Chinese University of Hong Kong

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

©2010 Irwin King. All rights reserved.
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

<table>
<thead>
<tr>
<th>500</th>
<th>600</th>
<th>700</th>
<th>800</th>
<th>900</th>
<th>1000</th>
<th>1100</th>
<th>1200</th>
<th>1300</th>
<th>1400</th>
<th>1500</th>
<th>1600</th>
<th>1700</th>
<th>1800</th>
<th>1900</th>
<th>2000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Middle Ages</td>
<td>Late Middle Ages</td>
<td>Renaissance</td>
<td>Enlightenment</td>
<td>Age of Liberalism</td>
<td>Age of Revolution</td>
<td>World at War and Interwar Years</td>
<td>The Modern World</td>
<td>Medieval Age</td>
<td>High Middle Ages</td>
<td>The Reformation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A Brief History of the World

1750 Industrial Revolution
1945 ENIAC
1969 The MITS Altair
1975 Apple II
1981 Birth of Internet
1983 Time Magazine Person of the Year
1984 Apple Macintosh
1989 Birth of WWW
1996 Birth of XML
2004 Birth of Web 2.0
2006 Time Magazine Person of the Year

Attention Age
Social Networking

HOW TO USE WEB 2.0 IN THE ENTERPRISE

HAVEN'T YOU READ MY BLOG POST? I'VE FIRED YOU YESTERDAY

PART 1: COMMUNICATE WITH YOUR EMPLOYEES
Billionaires’ Shuffle

Facebook in 2004.02

at 23 and $1.5 billion later...

William Gates
Warren Buffett
Carlos Slim Helu & family
Mark Zuckerberg
William Gates
Carlos Slim Helu & family
Warren Buffett
Mark Zuckerberg
William Gates
Carlos Slim Helu & family
Warren Buffett
William Gates
Carlos Slim Helu & family
Warren Buffett
William Gates
Top 10 Most Populated Countries

as of July 2009

- China: 1,335
- India: 1,177
- United States: 308
- Indonesia: 231
- Brazil: 192
- Pakistan: 168
- Bangladesh: 162
- Nigeria: 154
- Russia: 141
- Japan: 127

 Millions
Top 10 Most Populated Countries

as of February 2010
Facebook’s Global Audience

Global Audience: 316,402,840

Data for 11/03/2009

United States
Country Audience: 94,748,820
Percent of Global Audience: 29.95%

United States Male / Female

United States Age Distribution

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

<table>
<thead>
<tr>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>More than 400 million active users</td>
</tr>
<tr>
<td>50% of our active users log on to Facebook in any given day</td>
</tr>
<tr>
<td>More than 35 million users update their status each day</td>
</tr>
<tr>
<td>More than 60 million status updates posted each day</td>
</tr>
<tr>
<td>More than 3 billion photos uploaded to the site each month</td>
</tr>
<tr>
<td>More than 5 billion pieces of content (web links, news stories, blog posts, notes, photo albums, etc.)</td>
</tr>
<tr>
<td>shared each week</td>
</tr>
</tbody>
</table>

### 10 Largest Countries

<table>
<thead>
<tr>
<th>1. United States</th>
<th>94,748,820</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. United Kingdom</td>
<td>22,261,080</td>
</tr>
<tr>
<td>3. Turkey</td>
<td>14,215,880</td>
</tr>
<tr>
<td>4. France</td>
<td>13,396,760</td>
</tr>
<tr>
<td>5. Canada</td>
<td>13,228,380</td>
</tr>
<tr>
<td>6. Italy</td>
<td>12,581,060</td>
</tr>
<tr>
<td>7. Indonesia</td>
<td>11,759,980</td>
</tr>
<tr>
<td>8. Spain</td>
<td>7,313,160</td>
</tr>
<tr>
<td>9. Australia</td>
<td>7,176,640</td>
</tr>
<tr>
<td>10. Philippines</td>
<td>6,991,040</td>
</tr>
</tbody>
</table>

### 10 Fastest Growing Over Past Week

<table>
<thead>
<tr>
<th>Country</th>
<th>Growth Rate</th>
<th>Growth</th>
<th>Current Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poland</td>
<td>12.46 %</td>
<td>137,900</td>
<td></td>
</tr>
<tr>
<td>Thailand</td>
<td>10.96 %</td>
<td>161,300</td>
<td></td>
</tr>
<tr>
<td>Portugal</td>
<td>9.81 %</td>
<td>80,040</td>
<td></td>
</tr>
<tr>
<td>South Africa</td>
<td>9.25 %</td>
<td>189,080</td>
<td></td>
</tr>
<tr>
<td>Taiwan</td>
<td>7.82 %</td>
<td>367,400</td>
<td></td>
</tr>
<tr>
<td>Romania</td>
<td>7.65 %</td>
<td>28,060</td>
<td></td>
</tr>
<tr>
<td>Germany</td>
<td>7.54 %</td>
<td>350,240</td>
<td></td>
</tr>
<tr>
<td>Malaysia</td>
<td>7.43 %</td>
<td>236,840</td>
<td></td>
</tr>
<tr>
<td>Indonesia</td>
<td>6.84 %</td>
<td>752,640</td>
<td></td>
</tr>
<tr>
<td>Iraq</td>
<td>6.72 %</td>
<td>6,380</td>
<td></td>
</tr>
</tbody>
</table>
# 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>QQ</td>
<td>Yahoo</td>
<td>FC2</td>
<td>Google</td>
<td>Orkut.br</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>YouTube</td>
<td>YouTube</td>
<td>Orkut.in</td>
<td>Universo Online</td>
<td>Facebook</td>
</tr>
<tr>
<td>5</td>
<td>Taobao</td>
<td>Myspace</td>
<td>Rakuten</td>
<td>YouTube</td>
<td>YouTube</td>
<td>Windows Live</td>
</tr>
<tr>
<td>6</td>
<td>163</td>
<td>MSN</td>
<td>Livedoor</td>
<td>Blogger</td>
<td>Globo</td>
<td>MSN</td>
</tr>
<tr>
<td>7</td>
<td>Google</td>
<td>Windows Live</td>
<td>Ameblo.jp</td>
<td>Rediff</td>
<td>MSN</td>
<td>Wikipedia</td>
</tr>
<tr>
<td>8</td>
<td>Sohu</td>
<td>Wikipedia</td>
<td>mixi</td>
<td>Facebook</td>
<td>Google</td>
<td>Blogger</td>
</tr>
<tr>
<td>9</td>
<td>Youku</td>
<td>Craigslist</td>
<td>Wikipedia</td>
<td>Wikipedia</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>Myspace</td>
</tr>
<tr>
<td>Alexa as of Oct 2010</td>
<td>China</td>
<td>USA</td>
<td>Japan</td>
<td>India</td>
<td>Brazil</td>
<td>Global</td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------</td>
<td>---------------</td>
<td>-----------</td>
<td>--------------</td>
<td>-------------</td>
<td>------------</td>
</tr>
<tr>
<td>1</td>
<td>Baidu</td>
<td>Google</td>
<td>Yahoo.jp</td>
<td>Google.in</td>
<td>Google.br</td>
<td>Google</td>
</tr>
<tr>
<td>2</td>
<td>QQ</td>
<td>Facebook</td>
<td>Google.jp</td>
<td>Google</td>
<td>Google</td>
<td>Facebook</td>
</tr>
<tr>
<td>3</td>
<td>Taobao</td>
<td>Yahoo!</td>
<td>FC2</td>
<td>Facebook</td>
<td>Orkut.br</td>
<td>YouTube</td>
</tr>
<tr>
<td>4</td>
<td>Sina</td>
<td>YouTube</td>
<td>YouTube</td>
<td>Yahoo!</td>
<td>YouTube</td>
<td>Yahoo</td>
</tr>
<tr>
<td>5</td>
<td>Google HK</td>
<td>Amazon</td>
<td>Ameblo.jp</td>
<td>YouTube</td>
<td>Universo Online</td>
<td>Windows Live</td>
</tr>
<tr>
<td>6</td>
<td>163</td>
<td>Wikipedia</td>
<td>Rakuten</td>
<td>Blogger</td>
<td>Windows Live</td>
<td>Baidu</td>
</tr>
<tr>
<td>7</td>
<td>Sohu</td>
<td>Twitter</td>
<td>Google</td>
<td>Orkut.in</td>
<td>Globo</td>
<td>Wikipedia</td>
</tr>
<tr>
<td>8</td>
<td>Google</td>
<td>eBay</td>
<td>Livedoor</td>
<td>Wikipedia</td>
<td>Blogger</td>
<td>Blogger</td>
</tr>
<tr>
<td>9</td>
<td>Soso</td>
<td>Craigslist</td>
<td>Wikipedia</td>
<td>Twitter</td>
<td>Yahoo!</td>
<td>Twitter</td>
</tr>
<tr>
<td>10</td>
<td>Youku</td>
<td>Blogger</td>
<td>mixi</td>
<td>Rediff</td>
<td>Terra</td>
<td>QQ</td>
</tr>
</tbody>
</table>
EU Commission on Social Computing

**Figure 2: The growth in active usage of social computing applications**

*Active internet users: “Thinking about using the internet, which of the following have you ever done?”*

- Watch video clips online
- Listen to live radio/audio online
- Visit a friend’s social network page
- Read blogs
- Manage a profile on a social network
- Create a profile on a social network
- Leave a comment on a blog site
- Upload my photos to a photo sharing site
- Start my own blog/weblog
- Upload a video clip to a video sharing site

[Ala-Mutka et al. 2009]

Source: (Universal McCann, 2009)
Figure 1: Adoption of Social Computing

EU Commission on Social Computing

[Ala-Mutka et al. 2009]
Topics in Social Computing

- Social Behavior Analysis and Modeling
- Social Media
- Social Network Theory and Models
- Link Analysis/Graph Mining/Large Graph Algorithms
- Learning to Rank
- Recommender Systems/Collaborative Filtering/Query Suggestion
- 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...
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**
The Brave New Words
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
Politics

The New York Times
Friday, June 19, 2009

The Lede
The New York Times News Blog
June 2, 2009, 7:05 PM

China’s Great Firewall Blocks Twitter
By ROBERT MACKEY

Search This Blog
Previous Post: Bloggers Ponder Last Message From Missing Jet’s Computer
Next Post: Punditry From Bin Laden and Zawahiri on Obama’s Trip to the Middle East

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

Catherine Henriette/Agence France-Presse — Getty Images
Commerce

- Social marketing
- Who are the brokers?
- Who can exert the most influence on buying/selling?
- How much should one advertise?
Public Health

- People's **behavior** can be monitored
- What is on people’s mind translates to **search queries**
- Google predicts flu trends...

**2007–2008 U.S. Flu Activity - Mid-Atlantic Region**

ILI percentage

4%

2%

0%
Twitter Pop Culture

- **Twisdom: Twitter Wisdom**

- **A Philosopher Ponders Life in 140 Characters or Less**
  - "I don’t know the key to success, but the key to failure is trying to please everybody." Bill Cosby Do what you know in your soul is right!
  - It is a miserable state of mind to have few things to desire, and many things to fear. – Francis Bacon

- **The Longest Poem In the World-the awesome twitter poem!** 956,644 verses this morning and ~4,000 a day!
The YouTube Generation
The Age of FaceBook
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
Social Bookmarking
Social News/Mash Up

On-line Games and Virtual Communities

Second Life is an online, 3D virtual world imagined and created by its Residents.

Social Entertainment

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.

Social Knowledge Sharing

Wikipedia

Knol™

Share what you know

Write and post a knol (nōl) — a unit of knowledge.

Search
searchable through popular search engines

Create
easy to write and manage

Control
each knol is owned by you, the author
Social/Human Computation
Human Computation
Games With A Purpose (GWAP)

- ESP
- Image tagging
- **Tag a Tune**
- Song tagging
- **Verbosity**
- Database of common knowledge description
- **Squigl**
- Image segmentation
- **Matchin**
- Image search by aesthetic value
- **FlipIt**
- Finding similar image pairs
- **PopVideo**
- Video tagging
Crowdsourcing

Mechanical Turk is a marketplace for work.
We give businesses and developers access to an on-demand, scalable workforce. Workers select from thousands of tasks and work whenever it's convenient.

110,262 HITs available. View them now.

Make Money by working on HITs

HITs - Human Intelligence Tasks - are individual tasks that you work on. Find HITs now.

As a Mechanical Turk Worker you:
- Can work from home
- Choose your own work hours
- Get paid for doing good work

Find an interesting task

Work

Earn money

Get Results from Mechanical Turk Workers

Ask workers to complete HITs - Human Intelligence Tasks - and get results using Mechanical Turk. Register Now

As a Mechanical Turk Requester you:
- Have access to a global, on-demand, 24 x 7 workforce
- Get thousands of HITs completed in minutes
- Pay only when you're satisfied with the results

Fund your account

Load your tasks

Get results

Find HITs Now

Get Started
Foldit: Protein Folding Game

[Cooper et al, Nature 466, 756-760 (5 August 2010)]
Social Location-based Services
Web 2.0 Revolution

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

crew

teams

squad

organizations

cohorts

populations

markets

communities

partners

groups

binary

cardinal

integer

real

presence

identity

social role

reputation

expertise

trust

ownership

accountability

knowledge
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
On-Going Research

**Machine Learning**

- Smooth Optimization for Effective Multiple Kernel Learning (*AAAI’10*)
- Online Learning for Multi-Task Feature Selection (*CIKM’10*)
- Simple and Efficient Multiple Kernel Learning By Group Lasso (*ICML’10*)
- Online Learning for Group Lasso (*ICML’10*)
- 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

- Routing Questions to Appropriate Answerers in Community Question Answering Services (*CIKM’10*)
- Diversifying Query Suggestion Results (*AAAI’10*)
- 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**

- Recommender Systems with Social Regularization (*WSDM'11*)
- CMAP: Effective Fusion of Quality and Relevance for Multi-criteria Recommendation (*WSDM'11*)
- UserRec: A User Recommendation Framework in Social Tagging Systems (*AAAI'10*)
- Learning to Recommend with Social Trust Ensemble (*SIGIR’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**

- Collection of User Judgments on Spoken Dialog System with Crowdsourcing (*SLT’10*)
- 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*)
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
- **Service-based software platforms** and development **tools**
WSDM2011

Fourth ACM International Conference on Web Search and Data Mining

Home About Committee Authors Attendees Program Sponsors

Quick Links
• Workshop Information

Important Dates
Jul-26 Workshop-proposals
Aug-1 Paper-submission
Aug-23 Workshop-notification
Sep-1 Tutorial-proposals
Oct 15 Paper notification

Sponsors
Platinum Sponsors
Microsoft Research

WSDM (pronounced “wisdom”) is the premier international ACM conference covering research in the areas of search and data mining on the Web. The 4th ACM WSDM Conference will take place in Hong Kong, during February 9-12, 2011.

WSDM publishes original, high quality papers and presentations related to search and data mining on the Web and the Social Web, with an emphasis on practical but principled novel models of search, retrieval and data mining, algorithm design and analysis, economics implications, and in-depth experimental analysis of accuracy and performance.

Breaking News

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
"I don’t care about ‘Six Degrees of Kevin Bacon.’
I’m one degree away from a pork chop."
Q & A