Introduction to Collaborative Web

Sergej Zerr
zerr@L3S.de
Outline

• Collaborative Advantages
  • Wisdom of crowds
  • Conditions for a successful collaboration

• Using Collaborative Data
  • Gathering Data from Social Web / Mechanical Turk
  • Inter Rater Agreement

• Search and Sensemaking processes, the overview
  • Collaboration opportunities in (Web) search

• Overview Papers
  • Collaborative Search, Search result relevance judgments,..

• Small experiment
  • Can we collaborate?

• Discussion
Collaboration

Often we need more than one hand

Sometimes more than one brain

“Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies and Nations”

James Suroewicki
Direct Collaboration in WWW can be used:

- Collaborative tagging,
- Favorite assignments,
- Click logs,
- Data gathering,
- Recommendations,
- ect., ect., ect. . .

Tags: Rainbow, Sea, Island, Green, Palmtree, Maui
Social Computer

![Diagram showing utility vs. number of contributors. The x-axis represents the number of contributors ranging from 10 to 10,000+. The y-axis represents utility. The curve shows that as the number of contributors increases, the utility decreases, reaching an equivalent or greater utility under the curve. There are images of Homer Simpson and other Simpson characters to illustrate the concept.]

Equivalent, or greater, utility under the curve
Social Computer

Utility

Expert $$$$$

Masses $

4,000 experts
80,000 articles
200 years to develop
Annual Updates

100,000 amateurs
1.6 Million articles
5 years to develop
Real-Time Updates

Britannica

Wikipedia

# of contributors

10

10,000+
Collaboration: Paradox

Collaborative work needs to be managed efficiently

Kasparov won against the world

http://spectrum.ieee.org/computing/hardware/multicore-is-bad-news-for-supercomputers
Collaboration

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diversity of opinion</td>
<td>Each person should have private information.</td>
</tr>
<tr>
<td>Independence</td>
<td>People's opinions aren't determined by the opinions of those around them.</td>
</tr>
<tr>
<td>Decentralization</td>
<td>People are able to specialize and draw on local knowledge.</td>
</tr>
<tr>
<td>Aggregation</td>
<td>Some mechanism exists for turning private judgments into a collective</td>
</tr>
</tbody>
</table>
Groupthink Symptoms:
Irving Lester Janis (26 May 1918 - 15 November 1990)

- Illusion of invulnerability
- Collective rationalization
- Belief in inherent morality
- Stereotyped views of out-groups
- Direct pressure on dissenters
- Self-censorship
- Illusion of unanimity
- Self-appointed ‘mindguards’

http://en.wikipedia.org/wiki/Groupthink
Collaboration

“The best collective decisions are the product of disagreement and contest, not consensus or compromise.”

“The best way for a group to be smart is for each person in it to think and act as independently as possible.”
Many Complementary Games

- ESP Game: label images
  → Image retrieval by text
- Squigl: match the labels to areas
  → Object recognition
- Matchin: find the better image
  → Image ranking
- FlipIt: memory with similar images
  → Near duplicate detection

- Other areas covered as well: label songs, find synonyms, describe videos
- See: www.gwap.com by Luis von Ahn
Re-Using Human Power

reCAPTCHA IS A FREE ANTI-BOT SERVICE THAT HELPS DIGITIZE BOOKS.

→ LEARN HOW reCAPTCHA WORKS

eBird

Report and track the birds online – any place, any time!
Outline

• **Collaborative Advantages**
  • Wisdom of crowds
  • Conditions for a successful collaboration

• **Using Collaborative Data**
  • Gathering Data from Social Web / Mechanical Turk
  • Inter Rater Agreement

• **Search and Sensemaking processes, the overview**
  • Collaboration opportunities in (Web) search

• **Overview Papers**
  • Collaborative Search, Search result relevance judgments,..

• **Small experiment**
  • Can we collaborate?

• **Discussion**
Social Computer

Human can (yet) solve some tasks more efficient and/or accurate as a machine would do.

- Captcha
- Classification (OCR)
- Image tagging
- Speech recognition
Social Computer – Good Question

„Mr. Burns saw Homer with the binoculars“

Matthew Lease and Omar Alonso: http://de.slideshare.net/mattlease/crowdsourcing-for-search-evaluation-and-socialalgorithmic-search
Social Computer – Good Interface

Throw the coin and tell us the result

• Head?
• Tail?

Results

• Head 61
• Tail 39

People often tend just to select the first option 🙁

Better: Some preliminary textual answer

• Coin type?
• Head or tail.

Matthew Lease and Omar Alonso: http://de.slideshare.net/mattlease/crowdsourcing-for-search-evaluation-and-socialalgorithmic-search
L3S Research – “What about image privacy?”

Private

Public

Work  Sea  Winter  Water

Sergej Zerr, Stefan Siersdorfer, Jonathon Hare, Elena Demidova. Privacy-Aware Image Classification and Search, SIGIR’12
Gathering average community notion of privacy

• We crawled “most recently uploaded” Flickr photos (2 Months)
• Started a social annotation game (over the course of 2 weeks)
• 81 users (colleagues, social networks, forum users), 6 teams

Points: 223

“Private are photos which have to do with the private sphere (like self portraits, family, friends, your home) or contain objects that you would not share with the entire world (like a private email). The rest is public. In case no decision can be made, the picture should be marked as undecidable.”

Sergej Zerr, Stefan Siersdorfer, Jonathon Hare, Elena Demidova Privacy-Aware Image Classification and Search, SIGIR'12
Sampling Data from Social Web: Tags

Tags Distribution of 2Mio Flickr "Streetart" Images

- 1-20 tags per photo
- >75 tags per photo
Sampling Data from Social Web: User activity

Simple random sample can result in a set dominated by few power user
Inter–Rater Reliability: What about ambiguity?

Nederland, netherlands, holland, dutch
Rotterdam, wielrennen, cycling, duck
le grand depart, tour de france,
Reklame, caravan, Funny Fotos

**Figure 2.** Ambiguous horse-and-figur used in Studies 2-4. From “Ambiguity of Form: Old and New.” by G. H. Fischer, 1968, Perception & Psychophysics, 4, p. 191. Copyright 1968 by the Psychonomic Society. Reprinted with permission.
Inter–Rater Reliability “Where is the cat?”

<table>
<thead>
<tr>
<th>Image 1</th>
<th>Image 2</th>
<th>Image 3</th>
<th>Image 4</th>
<th>Image 5</th>
<th>Image 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>V</td>
<td>V</td>
<td>X</td>
<td>V</td>
<td>V</td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>V</td>
<td>X</td>
<td>X</td>
<td>V</td>
<td>X</td>
<td>V</td>
</tr>
<tr>
<td>V</td>
<td>X</td>
<td>X</td>
<td>V</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

**Results**
Inter–Rater Reliability “Where is the cat?”

- Ideas
  - Introduce „Honeypot“ – a set of ground truth objects and select only good raters
  - Select only raters who rate close to average rating.
  - Other strategies?

<table>
<thead>
<tr>
<th>V</th>
<th>V</th>
<th>X</th>
<th>V</th>
<th>V</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>V</td>
<td>X</td>
<td>X</td>
<td>V</td>
<td>X</td>
<td>V</td>
</tr>
</tbody>
</table>

Results
Inter–Rater Reliability:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Naive approach: 3 cases out of 6 = 0.5 agreement

Kilem L. Gwet, Handbook of inter-rater reliability 2010
Inter–Rater Reliability: Cohen’s Kappa (1960)

Idea: We need to remove agreement achieved just by chance

$$\hat{\kappa} = \frac{p_a - p_e}{1 - p_e}$$

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th></th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>Yes</td>
<td>35</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>No</td>
<td>5</td>
<td>40</td>
</tr>
<tr>
<td>total</td>
<td>40</td>
<td>60</td>
<td>100</td>
</tr>
</tbody>
</table>

$$p_a = \frac{n_{11} + n_{22}}{n_{11} + n_{12} + n_{21} + n_{22}} = \frac{35 + 40}{100} = 0.75$$

$$p_e = \frac{55 \times 40 + 45 \times 60}{100 + 100 + 100 + 100} = 0.49$$

$$\hat{\kappa} \approx \frac{0.75 - 0.49}{1 - 0.49} = .51$$

Kilem L. Gwet, Handbook of inter-rater reliability 2010
Inter–Rater Reliability: Missing Values

Idea: Use partial ratings to estimate marginal probability only

\[
\hat{\gamma}_k = \frac{p_a - p_e}{1 - p_e}
\]

\[
p_a = \frac{n_{11} + n_{22}}{n - (n_{x1} + n_{x2} + n_{1x} + n_{2x})} = \frac{30 + 34}{100 - (5 + 8)} = .74
\]

\[
p_e = \frac{50 \times 40}{100} + \frac{42 \times 55}{100} = 0.431
\]

\[
\hat{\gamma}_k = \frac{0.74 - 0.431}{1 - 0.431} = .54
\]

Kilem L. Gwet, Handbook of inter-rater reliability 2010
Inter–Rater Reliability: Extentions

• **Multiple Raters/Categories:**
  • Fleiss 1971 – Average over random pairs of raters for random objects

• **Adjustment for Ordinal and Interval Data, Weighting:**
  • weight judgements using distances between categories.

• **Reduce Kappa Paradox:**
  • Split judgements into more certain / marginal.
  • Measures: $AC_1$, $AC_2$ (ordinal and interval data)

• **Check for statistical significance:**
  • The number of categories and/or raters matters.

Kilem L. Gwet, Handbook of inter-rater reliability 2010
Inter–Rater Reliability: Kappa Interpretations

<table>
<thead>
<tr>
<th>Kappa</th>
<th>Strength of Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 0.0</td>
<td>Poor</td>
</tr>
<tr>
<td>0.0 – 0.20</td>
<td>Slight</td>
</tr>
<tr>
<td>0.21 – 0.40</td>
<td>Fair</td>
</tr>
<tr>
<td>0.41 – 0.60</td>
<td>Moderate</td>
</tr>
<tr>
<td>0.61 – 0.80</td>
<td>Substantial</td>
</tr>
<tr>
<td>0.81 – 100</td>
<td>Almost Perfect</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Kappa</th>
<th>Strength of Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 0.0</td>
<td>Poor</td>
</tr>
<tr>
<td>0.0 – 0.40</td>
<td>Poor</td>
</tr>
<tr>
<td>0.41 – 0.75</td>
<td>Intermediate to Good</td>
</tr>
<tr>
<td>&gt;0.75</td>
<td>Excellent</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Kappa</th>
<th>Strength of Agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 0.20</td>
<td>Poor</td>
</tr>
<tr>
<td>0.21 – 0.40</td>
<td>Fair</td>
</tr>
<tr>
<td>0.41 – 0.60</td>
<td>Moderate</td>
</tr>
<tr>
<td>0.61 – 0.80</td>
<td>Good</td>
</tr>
<tr>
<td>0.81 – 100</td>
<td>Very Good</td>
</tr>
</tbody>
</table>

Please note: These interpretations were proven to be useful mostly in medical domain (diagnosis)

Kilem L. Gwet, Handbook of inter-rater reliability 2010
Useful human power for annotating the Web

- 5000 people playing simultaneously can label all images on Google in 30 days!
- Individual games in Yahoo! and MSN average over 5,000 players at a time

Possible contributions: attach labels to images in other languages, categorize web pages into topics
Outline

- **Collaborative Advantages**
  - Wisdom of crowds
  - Conditions for a successful collaboration

- **Using Collaborative Data**
  - Gathering Data from Social Web / Mechanical Turk
  - Inter Rater Agreement

- **Search and Sensemaking processes, the overview**
  - Collaboration opportunities in (Web) search

- **Overview Papers**
  - Collaborative Search, Search result relevance judgments,

- **Small experiment**
  - Can we collaborate?

- **Discussion**
What is relevant?

“Snow. Snow is relevant.”
Search and Sensemaking Process

Can collaboration improve the sensemaking process at any step?
Do you use collaborative search?
What are the typical collaborative search tasks?

- Watched over someone’s shoulder as he/she searched the Web, and suggested alternate query terms.
- E-mailed someone links to share the results of a Web search.
- E-mailed someone a textual summary to share the results of a Web search.
- Called someone on the phone to tell them about the results of a Web search.

Around 90% of Microsoft employees are engaged in collaborative search activities.

Collaborative Search:

- Query formulation
- Search process/browsing
- Save/Bookmark
- Annotate/Organize

How to support users in collaborative searching?
- Ideas
- Tools(Web 2.0)
Indirect Collaboration on Web:

Google uses search/click logs. For PageRank algorithm each link to a page serves as a vote for that page.

Amazon uses search/click logs. For item recommendation similar users are the indirect voters for the product.

ect., ect.
Outline

• Collaborative Advantages
  • Wisdom of crowds
  • Conditions for a successful collaboration

• Using Collaborative Data
  • Gathering Data from Social Web / Mechanical Turk
  • Inter Rater Agreement

• Search and Sensemaking processes, the overview
  • Collaboration opportunities in (Web) search

• Overview Papers
  • Collaborative Search, Search result relevance judgments,

• Small experiment
  • Can we collaborate?

• Discussion
Software support for Co-located search (CoSearch).

Software support for Co-located search (CoSearch).

<table>
<thead>
<tr>
<th>Qualitative criteria</th>
<th>5 point Likert scale (5=strongly agree)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Was CoSearch easy to use?</td>
<td>3</td>
</tr>
<tr>
<td>Were the colors useful?</td>
<td>4.5</td>
</tr>
<tr>
<td>Was the query queue useful?</td>
<td>4</td>
</tr>
<tr>
<td>Were color tabs useful?</td>
<td>4</td>
</tr>
<tr>
<td>Mobile view useful?</td>
<td>4</td>
</tr>
<tr>
<td>I would use CoSearch at work and in the school</td>
<td>4</td>
</tr>
<tr>
<td>I would use CoSearch at home</td>
<td>3.5</td>
</tr>
</tbody>
</table>
Spatial distributed Search (SearchTogether)

- (a) integrating messaging
- (b) query awareness
- (c) current results
- (d) recommendation queue
- (e)(f)(g) search buttons
- (h) page-specific metadata
- (i) toolbar
- (j) browser

Spatial distributed Search (SearchTogether)

<table>
<thead>
<tr>
<th>Qualitative criteria (5 points Likert scale)</th>
<th>Collective tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>SearchTogether helped to complete the joint task</td>
<td>3.9</td>
</tr>
<tr>
<td>SearchTogether is more effective than other tools</td>
<td>4.1</td>
</tr>
</tbody>
</table>

- 38% of all result lists were the consequence of using history
- 70 positive and 9 negative ratings
- 22 of 36 recommendation were viewed by the recipients

Improvement Through Collaborative Ranking

Choose Your Category of Expertise or Interest
Computers

Choose Your Query in the selected category to Rank Results
Java coding

Click

Rank

PLEASE RANK THE RESULTS BY CLICKING RANK FIRST

Java Programming Examples
Description: pdf • Javadoc HTML - Java Code Conventions (from Sun Microsystems) ... Source Code • Example Output - Logging Demo (1) - Java SDK 1.4 Logging API ...
URL: http://www.idevelopment.info/data/Programming/java/PROGRAMMING_Java_Programming.shtml

Ranking Options
Highly Relevant Relevant
Not Relevant Don't Know
This result is selected Now! Choose Your Ranking Options

Free Java Tutorials & Guide | Java programming source code
Description: Free java guide website to learn Java programming through examples. Beginner’s tutorials of plsql and sql with java source code.
URL: http://www.freejavaguide.com

WeSearch: Collaborative Sensemaking (Hardware support)

WeSearch: Collaborative Sensemaking

<table>
<thead>
<tr>
<th>Qualitative criteria (7 point Likert scale)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>The system was easy to use</td>
<td>6</td>
</tr>
<tr>
<td>High awareness of others' activity</td>
<td>5</td>
</tr>
<tr>
<td>Export record for later view was useful</td>
<td>6</td>
</tr>
<tr>
<td>Prefer clip cutting instead of clip search</td>
<td>6</td>
</tr>
</tbody>
</table>
Hardware support for Co-located (TeamSearch).

- Circles are categories: people, location, year, event

Hardware support for Co-located (TeamSearch).

<table>
<thead>
<tr>
<th>Qualitative criteria (7 points Likert scale)</th>
<th>Collective tokens</th>
<th>Parallel tokens</th>
</tr>
</thead>
<tbody>
<tr>
<td>I worked closely with the other members of my group to accomplish this task.</td>
<td>5.75</td>
<td>4.88</td>
</tr>
<tr>
<td>Members of the group communicated with each other effectively.</td>
<td>5.75</td>
<td>5</td>
</tr>
<tr>
<td>The group worked effectively as a team on this task.</td>
<td>5.75</td>
<td>4.81</td>
</tr>
</tbody>
</table>

Outline

• **Collaborative Advantages**
  • Wisdom of crowds
  • Conditions for a successful collaboration

• **Using Collaborative Data**
  • Gathering Data from Social Web / Mechanical Turk
  • Inter Rater Agreement

• **Search and Sensemaking processes, the overview**
  • Collaboration opportunities in (Web) search

• **Overview Papers**
  • Collaborative Search, Search result relevance judgments,..

• **Small experiment**
  • Can we collaborate?

• **Discussion**
Web Science – Investigating the Future of Information and Communication
Smalltalk, what should be the properties of a perfect collaborative system?