Human or Automatic Answers? A Users’ Based Study

Cristina González-Caro. University Pompeu Fabra, Barcelona
Mari-Carmen Marcos. University Pompeu Fabra, Barcelona
Liliana Calderón-Benavides. University Pompeu Fabra, Barcelona
Ricardo Baeza-Yates. Yahoo! Research, Barcelona

1. Why studying SEs and CQA systems?
2. Hypothesis
3. Types of user needs
4. Experimental setup
5. Results
6. Conclusions
7. Future work
1. Why studying SEs and CQA systems?

- Information need
- Search Engines
  - Index (stored information)
  - ranked web pages
- Communities Question-Answering
  - Users (adhoc information)
  - Index (stored information)
  - answers (data, explanations, web pages...)

Users

Index (stored information)

Communities Question-Answering

Information need

Search Engines

ranked web pages
1. Why studying SEs and CQA systems?

How can users take benefit from CQA as IR tools?

- For what kind of information needs?
- Does the previous knowledge of a topic make a difference?
- How does their knowledge area influence?

1. Why studying SEs and CQA systems?
2. **Hypothesis**
3. Types of user needs
4. Experimental setup
5. Results
6. Conclusions
7. Future work
2. Hypothesis

- **SEs** are better to ask for **objective** information
  - *Where Picasso was born?*
  - *How can I make a ‘caipirinha’?*

- **Communities QA** are better to ask for **subjective** information
  - *Which is the best beach to practice snorkel in Brazil?*

- **User background** and her **previous knowledge** influences what kind of information she prefers to obtain, and so, if they prefer information from SEs or CQA.
2. Hypothesis

- SE are better for asking for **objective** information
  - *Where Picasso was born?*
  - *How can I make an apple pie?*

- **Communities QA** are better for asking for **subjective** information
  - *Which is the best beach to practice snorkel in the Caribbean Sea?*

The **user background** and her **previous knowledge** influences what kind of information she prefers, and so, if they prefer information from SEs or CQA.

In order to understand these issues, we have developed a study with real users
3. Types of user needs

- **Kind** of information expected
  - Objective
  - Subjective

- **Cardinality**: number of possible answers
  - Just **one** right answer
  - **Many** possible correct answers

<table>
<thead>
<tr>
<th>Kind</th>
<th>Objective</th>
<th>Subjective</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cardinality</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Just one</td>
<td>OBJECTIVE</td>
<td>MULTI-ANSWERS</td>
</tr>
<tr>
<td>Many</td>
<td>OBJECTIVE</td>
<td>UNI-ANSWERS</td>
</tr>
</tbody>
</table>
13

How many kilobytes are in a terabyte?

What do I need to do to make PHP and MySQL work together?

How many kilobytes are in a terabyte?
1. Why studying SEs and CQA systems?
2. Hypothesis
3. Types of user needs
4. **Experimental setup**
5. Results
6. Conclusions
7. Future work
3. Experimental setup

- We have tested suitability:
  - With 84 users (60 + 24)
  - For 12 questions
  - In 6 information systems

1\textsuperscript{st} Choosing questions
2\textsuperscript{nd} Converting questions into queries
3\textsuperscript{rd} Submitting questions and queries
4\textsuperscript{th} Collecting answers and results
5\textsuperscript{th} Scoring the answers and suitability of results
1\textsuperscript{st} step. Choosing questions

- 12 Natural Language questions
- Topic: technologies
  
  - 4 objective and uni-answer
  - 4 objective and multi-answer
  - 4 subjective

2\textsuperscript{nd} step. Converting questions into queries

- 60 users did it for us
  
  - 20 from Inf. and Library Science
  - 20 from Social Sciences
  - 20 from Computer Science

- 60 users * 12 questions = 720 queries
3rd step. Submitting questions and queries

- 12 NL questions to 3 CQA
  - Yahoo! Answers
  - Live QnA
  - LinkedIn Answers

- 720 queries to 3 SE
  - Yahoo! Search
  - Google Search
  - Microsoft Live Search

4th step. Collecting answers and results

- From SEs
- From CQA
5th step. Scoring the answers/results

- **24 people** (3 backgrounds) evaluated relevance

They told us how much they knew about
- each topic
- every question
- Only **PDF files** to avoid users clicking links.

- The users did not know whether **answers came from** a SE, a CQA or another source.

---

- **Scores:**
  - Excellent: 100%
  - Enough: 66%
  - Poor: 33%
  - Not useful: 0%

**Question 1:**
Who was the Google creator and in what year did they create it?

**Web Pages**

<table>
<thead>
<tr>
<th>Web Page 1</th>
<th>Excellent</th>
<th>Enough</th>
<th>Poor</th>
<th>Not Useful</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Web Page 2</th>
<th>Excellent</th>
<th>Enough</th>
<th>Poor</th>
<th>Not Useful</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Web Page 3</th>
<th>Excellent</th>
<th>Enough</th>
<th>Poor</th>
<th>Not Useful</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Web Page 4</th>
<th>Excellent</th>
<th>Enough</th>
<th>Poor</th>
<th>Not Useful</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
5. Results

In which cases are CQA more useful than SEs?

A) Does the previous knowledge of users about topics or questions make a difference in how they evaluate the answers of CQA and SEs?

B) Do users knowledge area influences the users preference for one service or the other?

B) Does the type of information need influence the users preferences?

An ANOVA test was used to establish whether the differences between factors were significant
**A) NO. The previous knowledge of users about topics or questions did NOT make any difference in how users evaluated the answers**

<table>
<thead>
<tr>
<th>Query</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.213</td>
<td>0.809</td>
</tr>
<tr>
<td>2</td>
<td>0.063</td>
<td>0.802</td>
</tr>
<tr>
<td>3</td>
<td>0.567</td>
<td>0.454</td>
</tr>
<tr>
<td>4</td>
<td>1.006</td>
<td>0.377</td>
</tr>
<tr>
<td>5</td>
<td>0.300</td>
<td>0.586</td>
</tr>
<tr>
<td>6</td>
<td>1.693</td>
<td>0.190</td>
</tr>
<tr>
<td>7</td>
<td>0.553</td>
<td>0.459</td>
</tr>
<tr>
<td>8</td>
<td>0.128</td>
<td>0.723</td>
</tr>
<tr>
<td>9</td>
<td>3.143</td>
<td>0.050</td>
</tr>
<tr>
<td>10</td>
<td>0.589</td>
<td>0.463</td>
</tr>
<tr>
<td>11</td>
<td>0.091</td>
<td>0.940</td>
</tr>
<tr>
<td>12</td>
<td>3.207</td>
<td>0.045</td>
</tr>
</tbody>
</table>

**5. Results**

In which cases are CQA more useful than SE?

**A) Does the previous knowledge of users about topics or questions make a difference in how they evaluate the answers CQA and SE?**

**B) Does users’ knowledge area influence the users preference for one service or the other?**

**B) Does the type of information need influence the users preferences?**

An ANOVA test was used to establish whether the differences between factors are significant.
5. Results

In which cases are CQA more useful than SE?

A) Does the previous knowledge of users about topics or questions make a difference in how they evaluate the answers CQA and SE?

B) Do users knowledge area influences the users preference for one service or the other?

B) Does the type of information need influence the users’ preferences?

An ANOVA test was used to establish whether the differences between factors are significant.
C) **YES.** The type of information need influences their preference: CQA answered better to Objective Multi-Answer Questions

![Graph showing the relevance scores for different answer types](image)

- **F** → 9.512
- **p** → 0.002

1. Why studying SEs and CQA systems?
2. Hypothesis
3. Types of user needs
4. Experimental setup
5. Results
6. **Conclusions**
7. Future work
6. Conclusions

Search Engines

Communities Question-Answering

Index (stored information)

Index (stored information)

Users (adhoc information)

ranked web pages

answers (data, explanations, web pages...)

SEs can take benefit from CQA indexes, especially for some kind of questions and users
A big challenge: SEs should be able to detect

1. When users are expecting objective answers
2. When the right answers can be more than one
3. What is the users’ background

1. Why studying SEs and CQA systems?
2. Hypothesis
3. Types of user needs
4. Experimental setup
5. Results
6. Conclusions
7. Future work
6. Future work

Confirming these first experiment results
- With other topics, trivial topics included
- With real needs from query log analysis

Studying how to predict users expectations from their queries
- Understanding how people build queries according to their information needs

---

Thank you for your attention

Cristina González-Caro, cristina.gonzalez@upf.edu
Mari-Carmen Marcos, mcarmen.marcos@upf.edu
Liliana Calderón-Benavides, liliana.calderon@upf.edu
Ricardo Baeza-Yates, ricardo@baeza.cl