Crowdsourcing Interactions
Capturing query sessions through crowdsourcing

G. Zuccon*, T. Leelanupab*, S. Whiting*, E. Yilmaz§, J. M. Jose*, and L. Azzopardi*
*University of Glasgow, UK
§Microsoft Research, Cambridge, UK
{guido, kimm, stewh, jj, leif}@dcs.gla.ac.uk
eminey@microsoft.com
Query sessions in TREC

- *artificially generated reformulations*
  - Specification:
    1. Select a query from TREC Web Diversity Task
    2. Select one of the query’s subtopic
    3. Extract appropriate keywords
Query sessions in TREC

• artificially generated reformulations
  • Specification:
    1. Select a query from TREC Web Diversity Task
    2. Select one of the query’ subtopic
    3. Extract appropriate keywords

• relevance judgments made by NIST assessors
Alternative way?

- Is there an alternative way to capture sessions?
  - Well, let's get query logs from search engines!
  - But, they don’t give them to you 😞
  - So, what about crowdsourcing them?
What is Crowdsourcing?

Outsourcing a task to an undefined, generally large group of people

Crowdsourcing is the act of outsourcing tasks, traditionally performed by an employee or contractor, to an undefined, large group of people or community (a "crowd"), through an open call.
Use crowdsourcing for getting user interactions:

- **workers are asked to complete information seeking tasks** within a web-based crowdsourcing platform
- researchers can capture **logs of workers interactions**
- acquire entry and post-search information and statistics, to characterise user population.
- ~ similar to lab-based experiments?
1. Define Information Seeking Tasks
1. Define Information seeking task

Type the answer to our questions. Please, be as clear as possible.

Question 1: How many times was Moon a Pro Bowler?

Answer to question 1: _______________________

☐ I don’t know

Question 2: Who have coached Moon in professional football?

Answer to question 2: _______________________

☐ I don’t know

Question 3: List the professional teams for which Moon has been a player.

Answer to question 3: _______________________

☐ I don’t know
1. Define Information Seeking Tasks
2. Capture Interactions
2. Capture Interactions

Demo: http://research.alltheway.co.uk/?AssignmentId=123&HitId=123
2. Capture Interactions
2. Capture Interactions

Search the web for: Warren Moon

Page 1 of about 23,300,000 results (0.34 seconds)

Warren Moon - Wikipedia, the free encyclopedia
Harold Warren Moon (born November 18, 1956 in Los Angeles, California) is a former American professional gridiron football quarterback who played for the Canadian Football League's ...

Warren Moon: Biography from Answers.com
Warren Moon is a former professional gridiron football quarterback who played for the Canadian Football League's...

Hall of Famers » WARREN MOON
Warren Moon: Thank you very much. Wow. Just sitting up here listening to everyone speak, talk about their feelings about today is really giving me a heavy feeling inside me...

Warren Moon NFL & AFL Football Statistics - Pro-Football-Reference.com
Pos:QB, Career: 208 G, 204 TD (8th), 83.0 Rate (4th), 237.1 Yd/C (10th), 8xProBowl, Oilers/Vikings, 1984-2000, 2x Yds Leader, born in CA 1956

Warren Moon
Warren Moon was a quarterback in the National Football League playing most of his career with the Houston Oilers. Moon was also a quarterback with the Canadian Football League for...

Moon, Warren definition of Moon, Warren in the Free Online...
Moon, Warren (Harold Warren Moon, Jr.), 1956–, African-American football player, b. Los Angeles. Moon quarterbacked the Univ. of Washington Huskies to a Rose Bowl title in 1976...

Warren Moon and Tony Gonzales Bring their Legendary Pro Bowl... 
The National Football League's best and brightest will mix it up at Pearl Ultra Lounge in Waikiki on Thursday night at 10 p.m. for the Pro Bowl Player Party hosted...

Warren Moon Facts, information, pictures | Encyclopedia.com...
Warren Moon 1956– Football player. For seventeen seasons Warren Moon was a starting quarterback in the National Football League (NFL), most of that time for the Houston Oilers.

mark this page
It contains part of my answer.

marked!

mark this page
It contains part of my answer.

marked!
Crowdsourcing Interactions

1. Define Information Seeking Tasks
2. Capture Interactions
3. Acquire post-search information
3. Acquire post-search information

Answer the following questionnaire about your search experience (you will not be judged on the basis of your answer to this questionnaire). Please, complete the questionnaire ONLY AFTER your search.

The task you performed was (1=not at all; 5= completely):

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unclear</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>O</td>
</tr>
<tr>
<td>Easy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>O</td>
</tr>
<tr>
<td>Simple</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>O</td>
</tr>
<tr>
<td>Unfamiliar</td>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

I haven't known the answers before performing the task:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disagree</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>A</td>
</tr>
</tbody>
</table>

It was easy to complete the task:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disagree</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>A</td>
</tr>
</tbody>
</table>

I believe I have succeeded in the task:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disagree</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>A</td>
</tr>
</tbody>
</table>

Please provide any comments you may have below, we appreciate your input!
Crowdsourcing Interactions

1. Define Information Seeking Tasks
2. Capture Interactions
3. Acquire post-search information
4. Characterise User Population
4. Characterise user population

Require qualification test:

- **Quantitatively** characterise a user
- based on aptitude or Intelligence Quotient (IQ)
- estimate of whether workers are suitable for task
- group users
4. Characterise user population

Which word is closest in meaning to LIGAMENT?

1. Band
2. Nerve
3. Tendon
4. Appendage
5. Branch
Crowdsourcing Interactions: An example
Crowdsourcing Interactions: An example
Crowdsourcing Interactions: An example

![Graph showing query length over session queries for different sessions.](Image)
# Crowdsourcing Interactions: An example

<table>
<thead>
<tr>
<th># of interactions</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sessions with logged interactions</td>
<td>57</td>
</tr>
<tr>
<td>Sessions with 1 click</td>
<td>27</td>
</tr>
<tr>
<td>Sessions with 2 clicks</td>
<td>12</td>
</tr>
<tr>
<td>Sessions with 3 clicks</td>
<td>7</td>
</tr>
<tr>
<td>Sessions with 4 clicks</td>
<td>5</td>
</tr>
<tr>
<td>Sessions with 5 clicks</td>
<td>6</td>
</tr>
</tbody>
</table>

Total sessions: 266
Take Home Points

- Crowdsourcing **cheap** way to acquire (small) **query logs**
- **Heterogeneous** user population and searching behaviors
- To work on:
  - Assess/improve quality of captured data
  - How to get query specilisations? Generalisations?
Data is available

You can find the logs at:

http://www.dcs.gla.ac.uk/~guido/sir2011