Visual Search Engine Evaluation

Efthimis N. Efthimiadis
efthimis@u.washington.edu

Information School
University of Washington
Agenda

• Motivation
• Aims & Objectives
• Methodology
• Results
• Conclusions
Background & Motivation

• The presentation of information by SE is fundamental to their usability.

• Most SEs present a simple ranked list.

• Visualization of results is active & growing research area

• A number of SEs that use a visual or hybrid UI exist:  
  e.g., Kartoo, Quintura, Grokker, Mooter, WebBrain, PageBull, Snap

• few comparative usability studies of visual search engines
Comparative usability evaluation studies of visual search engines

- visually impaired users of Google, Yahoo, HotBot, Vivisimo & Kartoo (Andronico et al., 2004).
- visual vs. textual display of information using the Google, Yahoo, Grokker & Kartoo (da Silva & Freitas, 2006)
- treemap display of results was compared with Google, SRC, Kartoo & Clusty (Chu et al., 2007).

Comparative usability evaluation studies of SE are complicated (White & Marchionini, 2006; Käki & Aula, 2008):
- search is highly domain-dependent and user-dependent
- variation between individual’s information-seeking habits adds complexity to the selection of both appropriate experimental conditions & test metrics.
Aim

- investigate how people search when using a visual user interface (UI) compared to the traditional text-based UI.
Goals & Objectives:

(a) compare how people search using three types of UIs (visual, text, hybrid);

(b) identify user satisfaction level with each system for completing a search;

(c) establish how the visualizations help/hinder users with the search/navigation process;

(d) study how the query formulation and reformulation process is affected by the visual UIs.
METHODOLOGY
Methodology

• Participants
• Tasks
• Search engines
• Study design
• Data collection Instruments
Participants

<table>
<thead>
<tr>
<th></th>
<th>Undergraduate Students</th>
<th>Graduate Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 24</td>
<td>12 (4 male, 8 female)</td>
<td>12 (6 male, 6 female)</td>
</tr>
<tr>
<td>Age Range</td>
<td>18-20: 9</td>
<td>21-25: 2</td>
</tr>
<tr>
<td></td>
<td>21-25: 2</td>
<td>26-30: 5</td>
</tr>
<tr>
<td></td>
<td>46-50: 1</td>
<td>31-35: 3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>36-40: 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>41-45: 1</td>
</tr>
</tbody>
</table>

- Convenience sample
- Recruited from UW student population
- Grads: LIS and Information Management
- Undergrads: non-iSchool students taking an iSchool intro to tech class
Search Tasks

• Developed three scenarios
  (TREC interactive track, and Borlund & Ingwersen, 2000)

• Query Types used (Broder 2002):
  – Informational
  – Transactional
Scenarios

1. Aspirin
2. Browser security plug-in
3. Tour de France
Scenario 1 - Aspirin

• A family friend, a woman in her mid-50’s, is concerned about heart attacks because of her family history.

• She’s heard that taking a daily dose of aspirin can reduce the risk of heart attacks in some patients, but she has also heard that Aspirin has certain side effects.

• She will be visiting her doctor in a few days and wants to find information about this in order to have an educated conversation with her physician.

• To help her you volunteered to search and identify at least two pros or cons of taking large doses of Aspirin.
Scenario 2 – Security Plug-In

• You are concerned about protecting sensitive personal data and maintaining privacy when you use the internet.

• Plug-ins are small, free-downloadable programs that work with your Internet Browser such as Firefox or Internet Explorer.

• You use the Firefox web browser but you realize that there are more than 150 Firefox plug-ins related to privacy and security.

• Can you find authoritative information that recommends five to ten of the best plug-ins for your use?
Scenario 3 – Tour de France

• You live in Seattle, WA.
• Next summer you’re planning to take a bicycle trip that will follow the route of the annual “Tour De France” bike race.
• You’re interested in several pieces of information.
  1. How much will it cost to ship your bicycle to Paris?
  2. Identify some package tours that might include lodging, ground transportation, a guide and the price ranges.
  3. Find narratives and tips from people who have made this trip in the past.
SEARCH ENGINES
STUDY DESIGN
Study Design

• within-subjects design

• Latin square design to account for learning effects for both systems and topics

User {u#}, System {g, k, q}, Scenario {1, 2, 3}

<table>
<thead>
<tr>
<th></th>
<th>u1</th>
<th>g1</th>
<th>k2</th>
<th>q3</th>
<th>u4</th>
<th>k1</th>
<th>q2</th>
<th>g3</th>
</tr>
</thead>
<tbody>
<tr>
<td>u2</td>
<td>g2</td>
<td>k3</td>
<td>q1</td>
<td></td>
<td>u5</td>
<td>k2</td>
<td>q3</td>
<td>g1</td>
</tr>
<tr>
<td>u3</td>
<td>g3</td>
<td>k1</td>
<td>q2</td>
<td></td>
<td>u6</td>
<td>k3</td>
<td>q1</td>
<td>g2</td>
</tr>
</tbody>
</table>

• no user training on SE prior to study
  – assumption Web SE need to be intuitive and no training should be required to search (Hearst 2007)
DATA COLLECTION INSTRUMENTS
Questionnaires

• Data was gathered using Catalyst, a web-based survey tool
  – questions on a 9-point scale

• Initial survey (x1);
  – gathered demographic data and information about the subject’s computer use, web search experience, history & habits

• Pre-search survey (x3);
  – gathered information on the subject’s prior knowledge of the scenario, after reading the scenario

• Post-search survey (x3);
  – gathered information about the subject’s experience with a search engine and scenario

• Final survey (x1);
  – gather comparative information about the three search engines & search process
Recording user/SE interaction

DejaClick plug-in for the Firefox browser.

• Records entire session with Audio & Video, allows playback

• Collects information on:
  – keywords used to search,
  – URLs visited,
  – time it takes to complete the search task.
1. Begin recording

2. Indicates that recording is on

3. Click green button to stop recording

4. Save xml file of recording here

University of Geneva, March 18
2010
Overview of the

SEARCH ENGINE USER INTERFACES
KartOO Results Page

White text are suggested keywords for search refinement; thumbnails are sites that match the current search.
Hovering the cursor over a keyword adds it to the search; clicking on keyword performs new search (next slide).

Lines connect to sites that match the new search.
KartOO Results Page: Using keyword refinement

Search with added keyword
Hovering over thumbnail brings up a capsule of the site in the left nav bar. Relevant keywords are also highlighted. Clicking on thumbnail opens the page in a new window.
Quintura Search Results

Default setting uses 2/3 of page for Google-like list of search results

Search terms are highlighted with blue text. Suggested search refinements are in black or gray, depending on expected relevance.
Quintura: Using the term cloud for search refinement

Hovering over a term adds it to the query. The list is also updated to provide a preview of the results. However, this is slow and often confused users.
Quintura: Using the term cloud for search refinement

Clicking back arrow takes you back to the previous search

Clicking on term performs search with added keyword
All three search engines provide correct answers with the initial search. KartOO confuses users – who don’t realize they’ve answered the question! Quintura less so.
• Frequency & Satisfaction of web search
• Interaction time
• # of queries
• # of terms
• Familiarity with scenario topic
• % of correct responses
• Perceived satisfaction w/results & SE
• Testimonials
• Would you search Kartoo/Quintura again?

Selective presentation of

RESULTS
Frequency & satisfaction with web search (self-reported)

- Scale on y-axis: 1 (no satisfaction/low frq) – 9 (high satisfaction)
- Frequency: Grad (3) > Undergrad (1)
- Satisfaction: Grad (4) < Undergrad (2)
Average Interaction Time per SE, Scenario, and Group (in Minutes)

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Undergraduates</th>
<th>Graduates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Google</td>
<td>5.4</td>
<td>7.3</td>
</tr>
<tr>
<td>Kartoo</td>
<td>4.5</td>
<td>6.7</td>
</tr>
<tr>
<td>Quintura</td>
<td>5.3</td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>6.2</td>
<td></td>
</tr>
<tr>
<td>s1</td>
<td>3.4</td>
<td>4.4</td>
</tr>
<tr>
<td>s2</td>
<td>4.4</td>
<td>6.2</td>
</tr>
<tr>
<td>s3</td>
<td>3.1</td>
<td></td>
</tr>
<tr>
<td>Tour</td>
<td>7.9</td>
<td>7.8</td>
</tr>
</tbody>
</table>

University of Geneva, March 18 2010
Queries: selected examples

Average number of query terms per scenario:
\[ s1 = 5; \ s2 = 4; \ s3 = 3.7 \]

S1

- aspirin (single word)
- aspirin heart attack
- aspirin "large dose" "pros and cons"
- "heart attacks" aspirin site:.gov
- pros and cons of taking Aspirin for heart attacks (natural language)
- "heart attacks" aspirin ("daily dose"|dosage) "side effects" pros cons inurl:webmd (advanced)

S2

- best plug-ins
- Firefox plug-ins
- firefox privacy plug in
- best security plug-ins for web browsers
- firefox plugins security recommendation
- tour de france
- paris
- Bicycle shipping cost
- bicycle shipping seattle paris
- USPS bulk shipping
- france transportation cost info
- Price to ship bicycles from seattle to paris
- Bicycle tours, "Tour De France"
- tips for people who have traveled to Paris for Tour De France
- packages tours include lodging, ground transportation, guides and price ranges in Paris
Average number of queries

- Overall
- By scenario
- By search engine
• Familiarity with the subject matter (9=high)
  – 1 (aspirin use) =  5.3
  – 2 (privacy plug-ins) =  4.3
  – 3 (tour de France) =  4.1

• average difficulty per SE and scenario
  (9=most difficult)
Correctness of results by scenario & by subject group

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Undergraduates</th>
<th>Graduate Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, Aspirin - pros</td>
<td>58%</td>
<td>55%</td>
</tr>
<tr>
<td>1, Aspirin - cons</td>
<td>33%</td>
<td>45%</td>
</tr>
<tr>
<td>2, Plug-in</td>
<td>8%</td>
<td>36%</td>
</tr>
<tr>
<td>3, shipping</td>
<td>8%</td>
<td>18%</td>
</tr>
<tr>
<td>3, tours</td>
<td>58%</td>
<td>27%</td>
</tr>
<tr>
<td>3, narratives</td>
<td>33%</td>
<td>18%</td>
</tr>
</tbody>
</table>
median perceived satisfaction with the search results and with the search engine

<table>
<thead>
<tr>
<th>Perceived Satisfaction</th>
<th>Undergraduates</th>
<th>Graduate Students</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Results</td>
<td>SE</td>
</tr>
<tr>
<td>1, Aspirin, google</td>
<td>8.5</td>
<td>9.0</td>
</tr>
<tr>
<td>1, Aspirin, kartOO</td>
<td>4.5</td>
<td>3.5</td>
</tr>
<tr>
<td>1, Aspirin, quintura</td>
<td>7.5</td>
<td>8.5</td>
</tr>
<tr>
<td>2, plug-in, google</td>
<td>6.5</td>
<td>7.0</td>
</tr>
<tr>
<td>2, plug-in, kartOO</td>
<td>8.0</td>
<td>3.5</td>
</tr>
<tr>
<td>2, plug-in, quintura</td>
<td>7.0</td>
<td>6.0</td>
</tr>
<tr>
<td>3, Tour de F, google</td>
<td>4.5</td>
<td>7.0</td>
</tr>
<tr>
<td>3, Tour de F, kartOO</td>
<td>3.0</td>
<td>2.0</td>
</tr>
<tr>
<td>3, Tour de F, quintura</td>
<td>6.5</td>
<td>7.0</td>
</tr>
</tbody>
</table>

9 point scale, with 1 being not at all satisfied, and 9 being completely satisfied
Do you have any additional comments regarding the search engine?

KartOO

• It doesn't give me any clear indication of actual relevance or the real content of the page to which I would be navigating.

• I'm sorry to say that the "visual" aspect of this search interface doesn't seem helpful...

• It gave me a headache.

• The layout is terrible. You have to search manually through the search results. Not good.

• ...who made that?

• The layout was significantly different than what I'm used to seeing with a search engine. That threw me off for a second, but it was also relatively easy to adjust to the "new" layout style. I liked that the layout was based on graphics, rather than simply text.

• I liked the amoebas. Also, I thought the groupings were helpful in choosing sites.
Do you have any additional comments regarding the search engine?

Quintura

- I did not like it at all. Visually confusing and a ridiculous waste of technology.

- Aside from the cloud on the left, I did not find it noticeably different from Google.

- "cloud" idea is interesting, but no one will use it if the google way is right there.

- The right side of the search engine was similar to google, therefore I felt used to it. the left side was completely difficult.

- I preferred the dual display of top hits on the right and the graphic display of subject relationships on the left. On the right I had the option of getting directly to sites. On the left, it supported serendipitous exposure to related topics. If I had difficulty articulating my search question, this contextualization of related topics could assist in refining how I phrased my terms.
Do you have any additional comments regarding the search engine?

Google

• Even though the search didn't go well, I was still satisfied with it.

• Google is my favorite search engine.

• Typically, I'm very satisfied with Google, but in this case, it was very frustrating.
Would you use K or Q again?

- **Quintura**
  - Grad: 67% Yes, 25% No, 8% Don't know
  - Undergrad: 67% Yes, 17% No, 17% Don't know

- **KartOO**
  - Grad: 25% Yes, 67% No, 8% Don't know
  - Undergrad: 17% Yes, 67% No, 17% Don't know
participants had most difficulty with Kartoo,

they felt at home when searching with Google’s text-based UI

they opted to use the text-UI of Quintura because of familiarity, and either ignored or criticized the term cloud.

CONCLUDING REMARKS
Visual Search: summary of comments

Searchers had:

• Difficulties interacting with graphical UIs

• Difficulties in understanding what is displayed & how to relate the “visualization” to the “text search” they are familiar with

• Increased frustration with search interaction & search results
Questions ???