Effective summarisation for search engines

Ph.D. Research

Lorena Leal-Bando
Supervisors: Andrew Turpin and Falk Scholer

School of Computer Science & Information Technologies
Royal Melbourne Institute of Technology
Melbourne, Australia

lorena.lealbando@student.rmit.edu.au
{andrew.turpin,falk.scholer} @ rmit.edu.au

May 29, 2009
Information everywhere

1 Original figure in [4]
Search engines results

**Rodrigo y Gabriela**
This form will take you to the Rodrigo y Gabriela mailing list website, and add you to the official Rodrigo y Gabriela mailing list...

*Tour Dates - Watch - History - Shop*
www.rodgab.com/history.htm - 27k - Cached - Similar pages

**Rodrigo Y Gabriela**
Satori - Experiencing self doubt shortly after arriving in Ireland, Rodrigo y Gabriela were at a
time ebb. No money, poor English and a deep sense of...
www.rodgab.com/history.htm - 27k - Cached - Similar pages

More results from www.rodgab.com:

**Stairway to Heaven live (Rodrigo y Gabriela)**
4 min 37 sec - 15 Jun 2008
www.youtube.com Rodrigo y Gabriela stairway to heaven live in The Olympia
www.youtube.com/watch?v=VNz5o9TU050

**RMIT UNIVERSITY**

Search Entire RMIT website for Induction course Gri

- Site map
  - Home
  - A-Z index
  - About RMIT
  - Our Organisation
  - Study at RMIT
  - Current Students
  - Industry and Business
  - Industry engagement
  - Intersect
  - News and Events

**2009 Program**
- GRO Workshops for HDR Candidates: 2009 Program ... that contain various resources and links to courses you can do online. Some members of RMIT's ... version of the online Higher Degrees by Research Induction course at RMIT, which is recommended to all new...
- Description: GRO Workshops for HDR Candidates; 2009 Program Summary
- Our Organisation / Research / Graduate Research / Current Research Students / Candidature Basics / 2009 Program

- Basic HDR Candidature Information...
- 2column true true hdr/basics Inger Mewburn Induction course > Quick candidature Information guide > ... please contact the Graduate Research Office on gra@rmit.edu.au. Top of page Quick
- Description: Essential Higher Degrees by Research Candidature Information
- Our Organisation / Research / Graduate Research / Current Research Students / Candidature Basics

- Path
- Title
Different type of information needs

<table>
<thead>
<tr>
<th>Query</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Navigational</td>
<td>Find the homepage of RMIT University</td>
</tr>
<tr>
<td>Transactional</td>
<td>Online Gaming</td>
</tr>
<tr>
<td>Informational</td>
<td>How do hurricanes get their names?</td>
</tr>
</tbody>
</table>

- According to Jansen et al. [5] more than 80% of Web queries involve an *informational task*.
Objectives

- We will investigate how to effectively summarise documents in order to present coherent and useful information for informational queries.
- Study user behaviour while they read
  - What parts of a document do users focus their attention on?
  - How do users compose their own summaries?
  - How does the document type affect users’ criteria when creating a summary?
We will investigate how to effectively summarise documents in order to present coherent and useful information for informational queries.

Study user behaviour while they read
- What parts of a document do users focus their attention on?
- How do users compose their own summaries?
- How does the document type affect users’ criteria when creating a summary?

Propose an algorithm to improve the generation of snippets for search engines.
Summaries

- Abbreviated versions of a document.
- Extracting highly ranked sentences [6], [7].
- Combining sentences from different documents [3], [8].
Interactions with search engines results

- Surveys: Asking directly to users about their preferences.
- Log files: web site’s name, query, mouse activity [1].
- Eye tracker: monitors eye movements [2].
Experiment Design

- Ethical approval application
- Participants task
  - They will read an English document on the screen.
  - They will type a short summary of the document according to a given query (2 queries per document).
  - Then, they will select any part from the document to build a summary.
  - Finally, answer a survey.
Eye Tracker

What does Tobii T60 do?

Features:
- Free head movements
- Calibration process
- Fully automatic tracking and precision timing gaze data
- Tolerance to varying light conditions
To detail user studies (type of document)
Training phase with Tobii T60
Analysis result
Questions?
Thanks for your attention!
C. Clarke, E. Agichtein, S. Dumais, and R. White.
The influence of caption features on clickthrough patterns in web search.

E. Cutrell and Z. Guan.
What are you looking for?: an eye-tracking study of information usage in web search.

K. Filippova and M. Strube.
Sentence fusion via dependency graph compression.

D. Hawking.
Challenges in enterprise search.

B. J. Jansen, D. Booth, and A. Spink.
Determining the user intent of web search engine queries.

H. P. Luhn.
The automatic creation of literature abstracts.
A. Tombros and M. Sanderson.  
Advantages of query biased summaries in information retrieval.  

I. Varlamis and S. Stamou.  
Semantically driven snippet selection for supporting focused web searches.  