A Taxonomy of Web Search

by Andrei Broder
Overview

- Motivation
- Classic model for IR
- Web-specific Needs
- Taxonomy of Web Search
- Evaluation
- Evolution of Search Engines
- Conclusions
Motivation

- central tenet – information need

But: show me url of the site where I can find a map of New York
I want to buy a new computer

intend behind web search not always informational
web-specific needs should be taken into account
Aims of this Paper

Aims:

- point out the difference between classic IR and web search
- introduce and analyze a taxonomy of web searches
- show how search engines deal with web-specific needs
Classic Model for IR

![Diagram showing the classic model for IR]

- Info need
- Query
- Corpus
- Matching Rules
- Query Refinement
- Results
Web-specific Needs

Task → Info need → Verbal form → Query → Query Refinement → Search Engine → Results

Corpus
A Taxonomy of Web Search

Classification of web queries into 3 categories:

1. **Informational**
   (acquire some information assumed to be present on one or more web pages)

2. **Navigational**
   (to reach a particular site)

3. **Transactional**
   (perform some web-mediated activity)
1. Informational Queries

**Intent:** acquire some information assumed to be present on one or more web pages

- information is in static form
- no further interaction is predicted

1. Where will WC 2018 be held?

   *WC 2018*

2. What is the current rank of Hannover 96?

   *table Hannover 96*
2. Navigational Queries

Intent: to reach a particular site

- user visited it in the past or assumes that it exists
- "known item" search or "home page finding task"
- only one right result

1. What is the official website of IBM?
   
   official website IBM

2. HP Store in Germany

   HP Germany
3. Transactional Queries

**Intent**: perform some web-mediated activity

- further interaction is expected

- main categories: shopping, finding servers, downloading various types of files

1. I need an accommodation in Rome

   *hotel Rome*

2. I want to buy a new bed

   *bed Ikea*
Scenario

Alice wants to buy a printer. – Transactional Query

Alice found 3 printers that she likes.
She wants more information about them. – Infomational Query

Alice wants some information about firma Lexmark.
She assumes that it has a site in Germany. – Navigational Query

Alice wants to download a document with a printer test. – Transactional Query
Evaluation

2 Methods:

- a survey of AltaVista users
  - presented to random users
  - users are self selected
  - a pop-up window with the questions

- analysis of the query log at AltaVista
to distinguish between navigational and non-navigational queries:

Which of the following describes best what you are trying to do?

- 24.53% I want to get to a specific website that I already have in mind
- 68.41% I want a good site on this topic, but I don’t have a specific site in mind
Survey Queries

➢ to distinguish between transactional and non-transactional queries:

Which of the following best describes why you conducted this search?

- 8.16% I am shopping for something to buy on the Internet
- 5.46% I am shopping for something to buy elsewhere than on the Internet
- 22.55% I want to download a file (e.g. music, images, programs, etc.)
- 57.19% None of these reasons
Survey Queries

➢ to distinguish between informational and non-informational queries:

Which of the following describes best what you are looking for?

14,83 % A site which is a collection of links to other sites regarding this topic

76,62 % The best site regarding this topic
a random set of 1000 queries from the daily AltaVista log

only English queries

sexually oriented queries are removed

queries that are neither navigational, nor transactional are assumed to be informational
## Summary

<table>
<thead>
<tr>
<th>Type of query</th>
<th>User Survey</th>
<th>Query Log Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Navigational</td>
<td>24.5%</td>
<td>20%</td>
</tr>
<tr>
<td>Informational</td>
<td>39%</td>
<td>48%</td>
</tr>
<tr>
<td>Transactional</td>
<td>36%</td>
<td>30%</td>
</tr>
</tbody>
</table>
Evolution of Search Engines

First generation (1995-1997)
- on-page data, close to classic IR, mostly informational queries
- AltaVista, Excite, WebCrawler, etc

- off-page, use of web-specific data such as link analysis, anchor-text, and click-through data, informational and navigational queries
- Google, DirectHit

Third generation (2000-now)
- attempt to ask the “need behind a query”, data from multiple sources, support for informational, navigational, transactional queries
Conclusion and Comments

- web search is task-driven
- search engines need to deal with different types of queries
- subtle difference between the navigational and informational queries
- some misleading in the evaluation part
Questions ???