A Taxonomy of Web Search

by Andrei Broder

Bahaeddin Eravci, Emre Yilmaz

2012

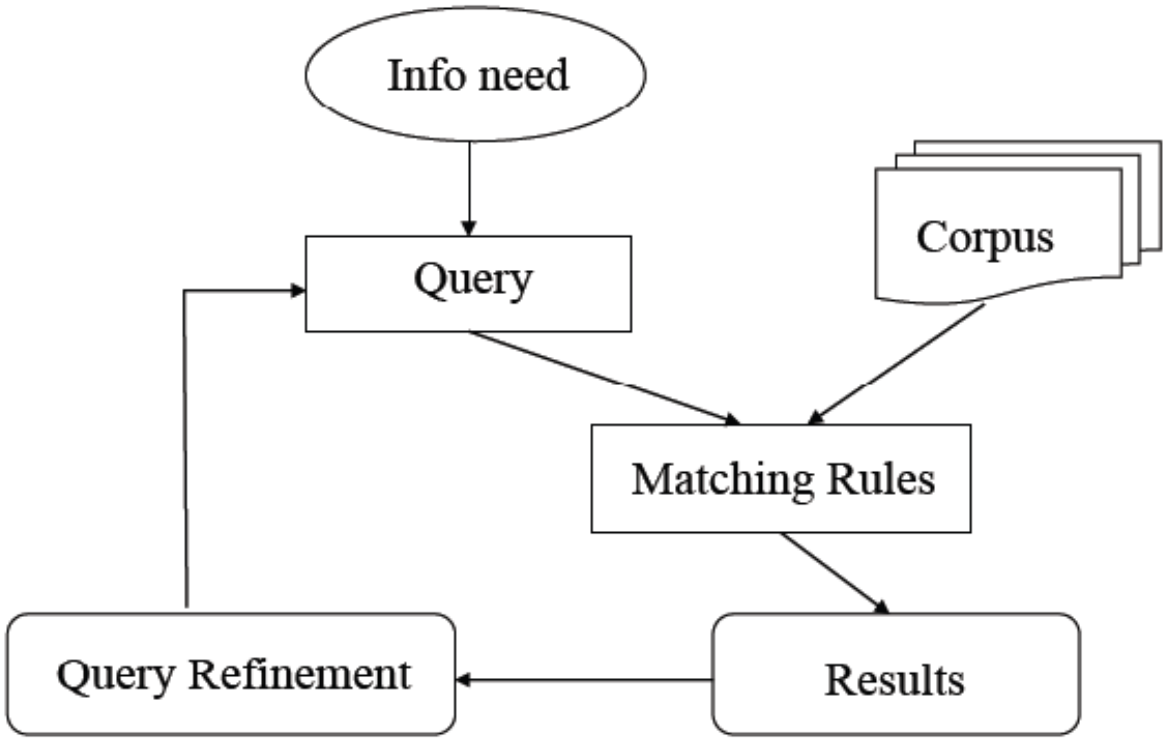
Outline

Outline

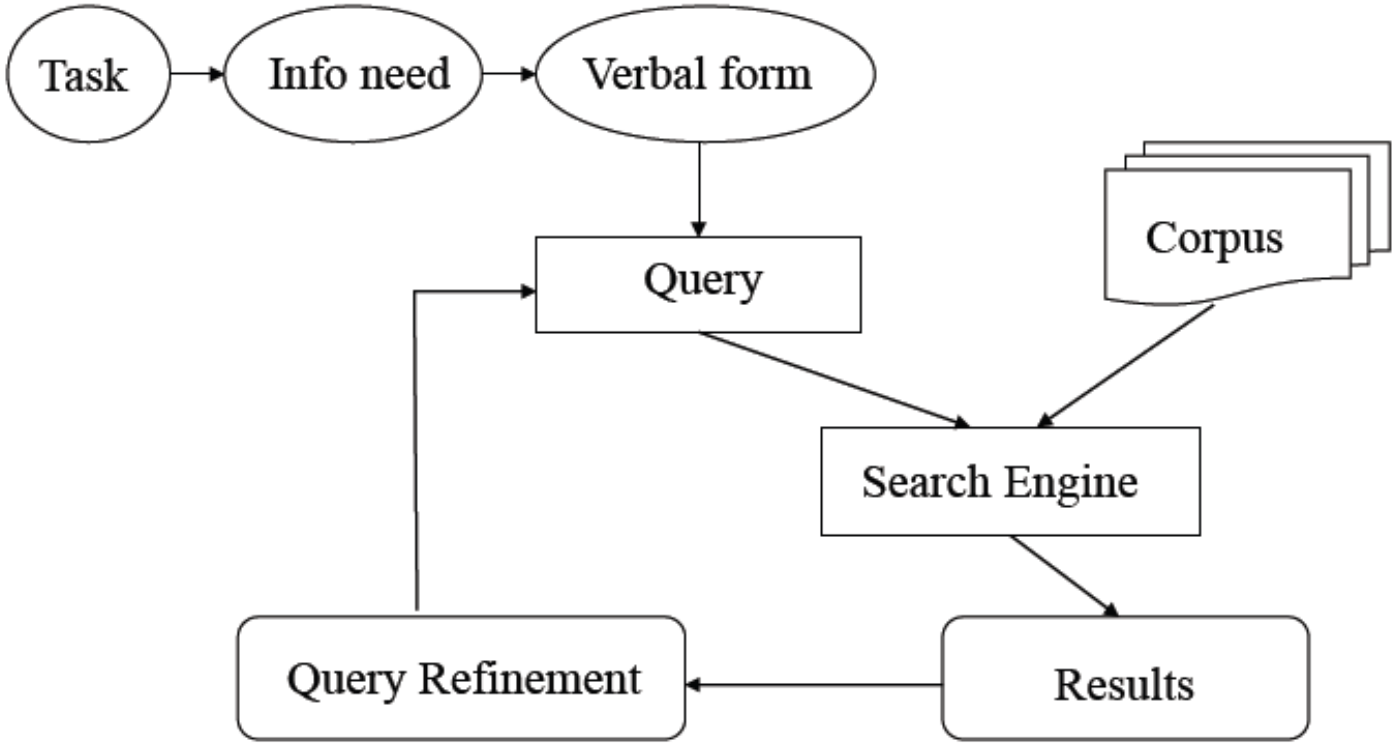
Aims of the Paper

- 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

The Classical Model for IR



Web-spesific Needs



Outline

Classification of Web Queries

1. Informational
2. Navigational
3. Transactional

Informational Queries

- Acquire some information assumed to be present on one or more web pages
- Information is in static form
- No further interaction is predicted

Example: Where will WC 2018 be held

```
WC 2018
```

Navigational Queries

- To reach a particular site
- User visited it in the past or assumes that it exists
- Only one right result

Example: What is the official website of IBM?

```
official website IBM
```

Transactional Queries

- Perform some web-mediated activity
- Further interaction is expected
- Main categories: shopping, finding servers, downloading various types of files

Example: I need an accommodation in Rome.

```
hotel Rome
```

Outline

User Survey

- A survey of AltaVista users
  - presented to random users
  - users are self selected
  - a pop-up window with the questions
- Questions to distinguish type of the query.

User Survey Questions

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

- I want to get to a specific website that I already have in mind
- I want a good site on this topic, but I don't have a specific site in mind

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

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

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

- A site which is a collection of links to other sites regarding this topic
- The best site regarding this topic

Log Analysis

- 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

Results

Table: Query Classification

Type of query	User Survey	Query Log Analysis
Navigational	24.5%	20%
Informational	39%	48%
Transactional	36%	30%

Outline

Evaluation of Search Engines

- **First generation (1995-1997)**
  - On-page data, close to classic IR, mostly informational queries
  - AltaVista, Excite, WebCrawler, etc.
- **Second generation (1998-1999)**
  - 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 (`San Francisco` : hotel reservation links, map server, weather server etc.)
  - Support for informational, navigational, transactional queries

Evaluation of Search Engines

- **First generation (1995-1997)**
  - On-page data, close to classic IR, mostly informational queries
  - AltaVista, Excite, WebCrawler, etc.
- **Second generation (1998-1999)**
  - 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 (`San Francisco` : hotel reservation links, map server, weather server etc.)
  - Support for informational, navigational, transactional queries

Evaluation of Search Engines

- **First generation (1995-1997)**
  - On-page data, close to classic IR, mostly informational queries
  - AltaVista, Excite, WebCrawler, etc.
- **Second generation (1998-1999)**
  - 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 (`San Francisco` : hotel reservation links, map server, weather server etc.)
  - Support for informational, navigational, transactional queries

Outline

Conclusion

- Web search is task-driven.
- Search engines need to deal with different types of queries.
- The main aim of third generation search engines is to deal efficiently with transactional queries via semantic analyses (understanding what the query is about) and blending of various external databases.

Questions