Refining a deployed system

- Once you have a system, with metrics, how do you consider changing the system to improve the metrics?
- A common approach is A/B testing.
  - This is done by Google for clients and Amazon for itself and probably many others.
- The idea:
  - Treat a small number of your users as experiments.
  - Have them use the different system.
  - Evaluate metrics on experimental group.
Evaluation in IR

- Gold standard approach

Diagram:
- Corpus
- Queries
- Gold Standard
- Standard Search Engine
- Proposed Alternative Search Engine
- Collect Metrics
- Collect Metrics
- Statistical Comparison of Metrics
- Experimental Validation or Reject of Alternative
Online A/B approach

- Requires
  - users
  - an infrastructure to support testing
  - metrics that don’t require a gold standard
Introducing Kindle: Amazon’s Revolutionary Wireless Reading Device

Amazon is excited to introduce Kindle—a wireless, portable reading device with instant access to more than 100,000 books, blogs, newspapers, and magazines. Whether you’re in bed or on the train, Kindle lets you think of a book and get it in less than a minute.

Learn more

Get Yourself a Little Something

See more in your Wish List

Save $50 on Select Toshiba Laptops

For a limited time save $50 on select Toshiba laptops offered by Amazon.com, with power Intel or AMD processors, large hard drives, and 13- or 15.4-inch display sizes. Hurry—savings end March 17.

19% off
Introducing Kindle: Amazon’s Revolutionary Wireless Reading Device

Amazon is excited to introduce Kindle—a wireless, portable reading device with instant access to more than 100,000 books, blogs, newspapers, and magazines. Whether you’re in bed or on the train, Kindle lets you think of a book and get it in less than a minute.

Learn more

Shop Irish Jewelry at Amazon.com

Claddagh
Clovers & Shamrocks
Celtic

Shop all Irish jewelry

Save $50 on Select Toshiba Laptops

For a limited time save $50 on select Toshiba laptops offered by Amazon.com, with power Intel or AMD processors, large hard drives, and 13- or 15.4-inch display sizes. Hurry—savings end March 17.
Website Optimizer (beta): Experiment List

Congratulations! You've signed up for Website Optimizer and are ready to create your experiment. As you move through the steps, your information will be saved.

+ Create a new experiment

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Status</th>
<th>Page Visitors</th>
<th>Conv.</th>
<th>Conv. Rate</th>
<th>Finish Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>No experiments</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Click Create experiment to get started.

My Client Center User Access [?]
Evaluation in IR

Snippets

• Little bits of text that summarize the page

[snippet]

• They function as an implicit tool for users to rank the results on their own (among those visible)

• The user does the final ranking

• Users are still biased by presented order though.
Snippets

• The goal of snippet generation is
  • present the most informative bit of a document in light of the query
  • present something which is self-contained
    • i.e., a clause or a sentence
  • present something short enough to fit in output
  • be fast, accurate (where are the snippets stored?)

• Challenges
  • Multiple occurrences of keyword in document
  • Poor English (or other language) grammar
Snippets

- Snippets can be **static**
  - A snippet for a web page is precompiled and always the same.

- Snippets can be **dynamic**
  - Depends on the query
    - “informatics”
    - “informatics definition”
Evaluation in IR

Snippets

- Snippets may contain
  - A few sentences from the web page
  - Meta data about the page
    - Author, Date, Title
  - Output of a text-summarization algorithm
    - Advanced technology that attempts to write snippets
  - Images from the document