On obtaining *Effort* based judgments for Information Retrieval

Manisha Verma (UCL)

Emine Yilmaz (UCL)

Nick Craswell (Microsoft)

Motivation

Relevance is the primary factor for ranking documents.

How do we determine what is **relevant**?

Train Judges

End Users

Smaller test collections
Supports repeated evaluation

Huge test collections
Suffers from different types
of noise

Motivation

Is batch evaluation == user-based evaluation?

NO, it has been repeatedly found (Hersh et al. SIGIR 2000, Maskari et al. SIGIR 2008) that these two evaluation mechanisms do not line up.

But WHY?

Motivation

- Current mechanism of judgment based evaluation does not take into account **'User effort'** (Yilmaz et al. CIKM 2014).

- A **judge can spend a lot of time** evaluating correctness of document for a given query.

- An **impatient user may not spend as much time** studying the document!

From Yilmaz *et al. CIKM 2014*

Contributions

- Collect effort based *judgments.*

- Determine *factors* associated with effort.

- Study their *association with user preferences*.

- Study retrieval performance when *effort and relevance* are taken into account.

User Model

1. When users first access the page, they ***quickly scan it*** to determine portions relevant to the query.

   FINDABILITY

2. This is ***followed by reading*** these paragraphs/snippets.

   READABILITY

3. Finally, user ***focusses on understanding*** these nuggets of information.

   UNDERSTANDABILITY

From Yilmaz *et al. CIKM 2014*

Methodology

- Collect **effort based (explicit) judgments** for each document for above parameters.

- Study user preferences.
  - **Control for relevance**: Collect user preferences with side-by-side comparison for documents of *same relevance grade*.
  - Analyse how these **preferences align with explicit judgments**?

Explicit Judgments

Factor Important for Satisfaction

Factor	p-val
Findability[+]	0.003
Readability[-]	0.364
Understandability[+]	0.054
Relevance[+]	0

Instructions

Suppose you submitted the following query to a search engine and document below was shown as result.

Search query: what are clouds

If page does not load please visit: http://www.weatherwizkids.com/weather-clouds.htm

Clouds

What are clouds?

A cloud is a large collection of very tiny droplets of water or ice crystals. The droplets are so small and light that they can float in the air.

How are clouds formed?

All air contains water, but near the ground it is usually in the form of an invisible gas called water vapor. When warm air rises, it expands and cools. Cool air can't hold as much water vapor as warm air, so some of the vapor condenses onto tiny pieces of dust that are floating in the air and forms a...

Would you be satisfied (happy) with this search result?
- Yes
- No
- Somewhat
- Can not judge (skip rest of the questions)

Is this document relevant to the query?
- Non Relevant
- Somewhat Relevant
- Relevant
- Highly Relevant

How difficult was it to understand the document?
- Very easy
- Easy
- Somewhat difficult
- Very difficult

Is the language easy to read?
- Very easy
- Easy
- Somewhat difficult
- Very difficult

Is it easy to find the answer of the query in the document?
- Very easy
- Easy
- Somewhat difficult
- Very difficult

There is More!

- Preference judgments also indicate that *Findability* is helpful in *distinguishing two equally relevant* documents.

- Comparison of top performing runs on TREC Web track datasets of 2012-2014 suggests that *performance of retrieval systems could be quite different when effort* (in our experiments Findability) is taken into account.

Thank you!