Sad horse

WebCLEF — Goodbye!

Maarten de Rijke

Agenda

- History
- WebCLEF this year
- What's next?

Where it all started . . .

Where it all started …

- So you want to do multilingual information access?

Where it all started . . .

- So you want to do multilingual information access?

- Work with web data, with web data in a European setting!

  - For many of us Europeans, dealing with multilingual text is all we know

A bit of history

- **Launched as a known-item search task in 2005, repeated in 2006**
  - **Resources created used for a number of purposes**

- **But there are information needs out there beside navigational ones, even on the web**

A bit of history

3. A taxonomy of web searches

In the web context the "need behind the query" is often not informational in nature. We classify web queries according to their intent into 3 classes:

1. **Navigational.** The immediate intent is to reach a particular site.
2. **Informational.** The intent is to acquire some information assumed to be present on one or more web pages.
3. **Transactional.** The intent is to perform some web-mediated activity.

Before we discuss these types in detail, we need to clarify that there is no assumption here that this intent can be inferred with any certitude from the query. The examples below might have alternative explanations

there beside navigational ones, even on the web

A bit of history

- **Launched as a known-item search task in 2005, repeated in 2006**
  - Resources created used for a number of purposes

- **But there are information needs out there beside navigational ones, even on the web**

A bit of history

- La... rch ta...

  - F... of p...

- Bu... s out th... even on...

Search Goal Hierarchy. Queries are only assigned to leaf nodes.
All examples are taken from actual AltaVista queries.

SEARCH GOAL	DESCRIPTION	EXAMPLES
1. Navigational	My goal is to go to specific known website that I already have in mind. The only reason I'm searching is that it's more convenient than typing the URL, or perhaps I don't know the URL.	aloha airlines duke university hospital kelly blue book
2. Informational	My goal is to learn something by reading or viewing web pages	
2.1 Directed	I want to learn something in particular about my topic	
2.1.1 Closed	I want to get an answer to a question that has a single, unambiguous answer.	what is a supercharger 2004 election dates
2.1.2 Open	I want to get an answer to an open-ended question, or one with unconstrained depth.	baseball death and injury why are metals shiny
2.2 Undirected	I want to learn anything/everything about my topic. A query for topic X might be interpreted as "tell me about X."	color blindness jfk jr
2.3 Advice	I want to get advice, ideas, suggestions, or instructions.	help quitting smoking walking with weights
2.4 Locate	My goal is to find out whether/where some real world service or product can be obtained	pella windows phone card
2.5 List	My goal is to get a list of plausible suggested web sites (I.e. the search result list itself), each of which might be candidates for helping me achieve some underlying, unspecified goal	travel amsterdam universities florida newspapers
3. Resource	My goal is to obtain a resource (not information) available on web pages	
3.1 Download	My goal is to download a resource that must be on my computer or other device to be useful	kazaa lite mame roms
3.2 Entertainment	My goal is to be entertained simply by viewing items available on the result page	xxx porno movie free live camera in l.a.
3.3 Interact	My goal is to interact with a resource using another program/service available on the web site I find	weather measure converter
3.4 Obtain	My goal is to obtain a resource that does not require a computer to use. I may print it out, but I can also just look at it on the screen. I'm not obtaining it to learn some information, but because I want to use the resource itself.	free jack o lantern patterns ellis island lesson plans house document no. 587

A bit of history

- Launched as a known-item search task in 2005, repeated in 2006
  - Resources created used for a number of purposes

- But there are information needs out there beside navigational ones, even on the web

Switching tasks

Switching tasks

- In 2007, launched a multilingual information synthesis task

    - "For a given topic, participating systems extract important snippets from web pages found using a web search engine"

Switching tasks

- In 2007, launched a multilingual information synthesis task

  - "For a given topic, participating systems extract important snippets from web pages found using a web search engine"

- Topics and assessments created by participants

  - Assessment environment provided by UAms

In 2007

- Only few participants
  - Afraid of the task?
  - Afraid of the amount of data to be processed?
  - Afraid of the content extraction task?
  - Afraid of the implementation effort?
  - Poorly advertized?
  - Just an odd year?

In 2008

- We gave participants the data
- We gave participants extracted content
- We gave participants a system
- We mailed, we announced, we phoned up, we invited, …

The outcome?

The outcome?

- First things first …
- The user model
  - knowledgable person writing a survey article on specific topic with clear goal and audience
  - user needs to locate items of information to be included in article and wants to use an automatic system for this purpose
  - user only uses online sources found via a Web search engine

The outcome (2)

- **Information needs specified as**
  - short topic title
  - free text description of goals and intended audience of the article
  - list of languages in which user is willing to accept information found
  - optional list of known sources: resources (URLs of web pages) considered relevant
  - optional list of Google retrieval queries

The outcome (3)

- topic title: *Paul Verhoeven*

  - description: *I'm looking for information on similarities, differences, connections, influences between Paul Verhoeven's movies of his Dutch period and his American period*

  - language: English, Dutch

  - known source(s):

    - http://en.wikipedia.org/wiki/Paul_Verhoeven, …

  - retrieval queries:

    - paul verhoeven (dutch AND american)}",
    - paul verhoeven (nederlandse AND amerikaanse OR hollywood OR VS)

The outcome (4)

- **61 multilingual topics**
  - 21 UK-ES
  - 21 EK-NL
  - 10 UK-RO-ES
  - 6 RU-EN
  - 2 EN-GE-NL
  - 1 RU-EN-NL

The outcome (5)

- **Test collection:**

  - Web docs found using Google with queries provided by topic providers
  - For each topic:
    - all "known" sources
    - top 100 hits (or less, depending on actual availability)—down from 1000 in 2007
    - for doc
      - URL, original content, plain text conversion (UTF-8), the query/queries that retrieved it, the rank it was returned

The outcome (6)

- ## Similar to, but simpler than in 2007

    - For a given topic all responses of all system were pooled into an anonymized randomized sequence of text segments
    - To limit amount of assessments required, for each topic only first 7,000 characters of each response were included (according to the ranking of the snippets in response)
    - For the pool created in this way for each topic, the assessors were asked to mark text spans that either (1) repeat the information already present in the known sources, or (2) contain new important information.
    - Unlike in 2007, assessors were not asked to group such text snippets into subtopics (by using nuggets), as the 2007 assessment results proved inconsistent with respect to nuggets.
    - The assessors used a GUI to mark character spans in the responses

So?

Runs and results

- **9 runs submitted by 3 research groups**
  - Twente, UNED, Salamanca
  - 1 run by best performing 2007 system
- **As of yesterday morning, 31 topics have been assessed, remainder expected by end of next week**

Runs and results

>	A	B	C	
1		P	R	
2	baseline	0.0978	0.0808	
3	ip2008	0.2537	0.2097	
4	ipt2008	0.2651	0.2191	
5	ipu2008	0.2218	0.1833	
6	Uned_RUN1	0.2720	0.2248	
7	Uned_RUN2	0.2243	0.1854	
8	Uned_RUN3	0.2445	0.2020	
9	usal_0	0.2334	0.1927	
10	usal_1	0.1299	0.1039	
11	usal_2	0.1032	0.0825	

Runs and results (2)

- Preliminary findings
  - runs consistently outperform baseline
  - there are some peaks, but there are no consistently "easy" topics
  - no run consistently outperforms other runs (unlike last year)

- This was WebCLEF's final year

- Parallel session tomorrow
  - 9.00–9.45
  - a look back & a look forward re: web ir

Let's step back

Why do you participate?

Why do you participate?

- **Because your funding agency tells you to do task-based evaluation?**

Why do you participate?

- **Because your funding agency tells you to do task-based evaluation?**

- **To create resources?**

Why do you participate?

- Because your funding agency tells you to do task-based evaluation?

- To create resources?

- To publish papers?
  - CLEF SIGIR, TOIS

Why do you participate?

- **Because your funding agency tells you to do task-based evaluation?**

- **To create resources?**

- **To publish papers?**
  - CLEF SIGIR, TOIS

- **To learn something?**

Why do you participate?

- **Because your funding agency tells you to do task-based evaluation?**

- **To create resources?**

- **To publish papers?**
  - CLEF SIGIR, TOIS

- **To learn something?**

- **Whatever it is, you need mass and innovation**

Oh yes, …

Oh yes, . . .

- **Mass and innovation will help you, your CV and your field**

Oh yes, . . .

- Mass and innovation will help you, your CV and your field

- But what if you could make a difference elsewhere?

- **Thanks!**