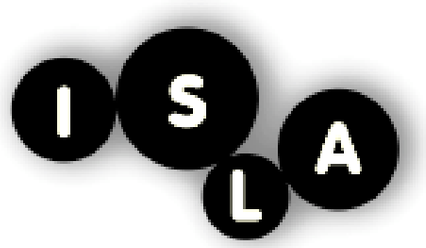
WebCLEF 2007 — The Overview

Valentin Jijkoun, Maarten de RIjke



Overview

Overview

- A bit of history

Overview

- A bit of history
- Task description

Overview

- A bit of history
- Task description
- Assessment

Overview

- A bit of history
- Task description
- Assessment
- Evaluation measures

Overview

- A bit of history
- Task description
- Assessment
- Evaluation measures
- Runs

Overview

- A bit of history
- Task description
- Assessment
- Evaluation measures
- Runs
- Results

Overview

- A bit of history
- Task description
- Assessment
- Evaluation measures
- Runs
- Results
- Conclusion

WebCLEF — A bit of history

WebCLEF — A bit of history

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

WebCLEF — 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

WebCLEF — A bit of history

- Launched as a known-item search task in 2005, repeated in 2006

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.

WebCLEF — 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

WebCLEF — A bit of history

- Launch
  - Resour

repeate

5,

- But the
  navigat

de

Table 1: The 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`

WebCLEF — 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

WebCLEF — 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

WebCLEF — 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

- WiQA
  - Pilot that ran at QA@CLEF 2006
  - Question answering using Wikipedia
  - Unidirected informational queries: “Tell me about X”

Task description

Task description

- Wishes
  - Task close to Real-World™ information need
  - Clear definition of a user
  - Multi-linguality should come naturally
  - Collections should be a natural source
  - Collections, topics, assessors' judgments re-usable
  - Challenging

Task description

- Wishes
  - Task close to Real-World™ information need
  - Clear definition of a user
  - Multi-linguality should come naturally
  - Collections should be a natural source
  - Collections, topics, assessors' judgments re-usable
  - Challenging

- Our hypothetical user
  - "A knowledgeable person, writing a survey or overview with a clear goal and audience in mind."
  - Locate items of information to be included in the article to be written, and use an automatic system to support this
  - Use online resources only

Task description (2)

Task description (2)

- User formulates her information need (“topic”)
  - A short **topic title** (e.g., title of the survey article)
  - A free text **description** of the goals and intended audience
  - A list of **languages** in which the user is willing to accept results
  - Optional list of **known source** (URLs of docs the user considers relevant)
  - Optional list of **Google retrieval queries**

Task description (2)

- User formulates her information need ("topic")
  - A short **topic title** (e.g., title of the survey article)
  - A free text **description** of the goals and intended audience
  - A list of **languages** in which the user is willing to accept results
  - Optional list of **known source** (URLs of docs the user considers relevant)
  - Optional list of **Google retrieval queries**
- Example
  - *title*: Significance testing
  - *description*: I want to write a survey (about 10 screens) or undergraduate students on statistical significance testing, with an overview of the ideas, common ideas and critiques. I will assume some basic knowledge of statistics
  - *language(s)*: English
  - *known sources*: http://en.wikipedia.org/wiki/Statistical_hypothesis_testing ..
  - *retrieval queries*: significance testing ; site:mathworld.wolfram.com ; ...

Task description (3)

Task description (3)

- Data
  - Close to Real-World™ scenario, but tractable
  - Define collection per topic
  - “Mashup”
    - All “known” sources specified
    - Top 1000 results per retrieval queries
    - Per result: query that retrieved it, rank, conversion (of HTML, PDF, PS) to plain txt

Task description (3)

- Data
  - Close to Real-World™ scenario, but tractable
  - Define collection per topic
  - “Mashup”
    - All “known” sources specified
    - Top 1000 results per retrieval queries
    - Per result: query that retrieved it, rank, conversion (of HTML, PDF, PS) to plain txt
- System’s response
  - Ranked list of plain txt snippets extracted from the sub-collection of the topic
  - Each indicates its origin

Assessment

Assessment

- Manual assessment by the topic creators
  - Somewhat similar to OTHER questions at TREC 2006
  - Blind
  - Pool responses of all systems into anonymized sequence of txt segments
  - For each response only include first 7,000 chars

Assessment

- Manual assessment by the topic creators
  - Somewhat similar to OTHER questions at TREC 2006
  - Blind
  - Pool responses of all systems into anonymized sequence of txt segments
  - For each response only include first 7,000 chars

- Asessor was asked …
  - To create a list of nuggets (“atomic facts”) that should be included in the article for the topic
  - Link character spans from a response to nugget
  - Different spans within a single snippet may be linked to multiple nuggets
  - Mark as “known” if a span expresses a fact present in known source

Responses for topic: iPhone opinions

Information need: *I want to compile a survey of people's first impressions about the new iPhone. What did they like? What do they complain about?*

Use your mouse to select some text from the results below. If the selection is ok you can press one of the buttons in the right column to mark the text as either covered in the known sources or belonging to one of the nugget.

Please select some text.

If the current selection contains information that is also in the known sources, you may mark the text as such using "Mark as known" button.

If you're happy with the current selection, you can add it to one an existing nuggets below, or you may create a new nugget.

Known sources

Assessing selected snippet

Mark as known

Add to new nugget

Responses for topic: iPhone opinions

Information need: *I want to compile a survey of people's first impressions about the new iPhone. What did they like? What do they complain about?*

Use your mouse to select some text from the results below. If the selection is ok you can press one of the buttons in the right column to mark the text as either covered in the known sources or belonging to one of the nugget.

Please select some text.

If the current selection contains information that is also in the known sources, you may mark the text as such using "Mark as known" button.

If you're happy with the current selection, you can add it to one an existing nuggets below, or you may create a new nugget.

Known sources

Assessing selected snippet

Mark as known

Add to new nugget

Responses for topic: iPhone opinions

Information need: *I want to compile a survey of people's first impressions about the new iPhone. What did they like? What do they complain about?*

Use your mouse to select some text from the results below. If the selection is ok you can press one of the buttons in the right column to mark the text as either covered in the known sources or belonging to one of the nugget.

Please select some text.

If the current selection contains information that is also in the known sources, you may mark the text as such using "Mark as known" button.

If you're happy with the current selection, you can add it to one an existing nuggets below, or you may create a new nugget.

Known sources

Assessing selected snippet

Mark as known

Add to new nugget

Responses for topic: iPhone opinions

Information need: *I want to compile a survey of people's first impressions about the new iPhone. What did they like? What do they complain about?*

Use your mouse to select some text from the results below. If the selection is ok you can press one of the buttons in the right column to mark the text as either covered in the known sources or belonging to one of the nugget.

Criticisms aside, the iPhone display is remarkable for its multitouch technology, which allows you to move your finger in a variety of ways to ma

We also found it somewhat tedious to scroll through long lists, such as the phone book or music playlists. Flicking your finger in an up or down motion will move you partway through a list, but you can't move directly to the bottom or top by swiping and holding your finger. On the other hand, the letters of the alphabet are displayed on the right side of the screen. By pressing a letter you can go directly to any songs or contacts beginning with that letter. But the lack of buttons requires a lot of tapping to move about the interface. For example, the Talk and End buttons are only displayed when the phone is in call mode. And since there are no dedicated Talk and End buttons, you must use a few taps to find these features. That also means you cannot just start dialing a number; you must open the dialpad first, which adds clicks to the process. The same goes for the music player: since there are no external buttons, you must call up the player interface to control your tunes. For some people, the switching back and forth may be a nonissue. But for mutlitaskers, it can grow wearisome.

Still, the interface and keyboard have a long way to go to achieve greatness. For starters, when typing an e-mail or text message the keyboard is displayed only when you hold the iPhone vertically. As a result,

Please select some text.

If the current selection contains information that is also in the known sources, you may mark the text as such using "Mark as known" button.

If you're happy with the current selection, you can add it to one an existing nuggets below, or you may create a new nugget.

Known sources

Assessing selected snippet

Mark as known

Add to new nugget

Responses for topic: iPhone opinions

Information need: *I want to compile a survey of people's first impressions about the new iPhone. What did they like? What do they complain about?*

Use your mouse to select some text from the results below. If the selection is ok you can press one of the buttons in the right column to mark the text as either covered in the known sources or belonging to one of the nugget.

Criticisms aside, the iPhone display is remarkable for its multitouch technology, which allows you to move your finger in a variety of ways to ma

We also found it somewhat tedious to scroll through long lists, such as the phone book or music playlists. Flicking your finger in an up or down motion will move you partway through a list, but you can't move directly to the bottom or top by swiping and holding your finger. On the other hand, the letters of the alphabet are displayed on the right side of the screen. By pressing a letter you can go directly to any songs or contacts beginning with that letter. But the lack of buttons requires a lot of tapping to move about the interface. For example, the Talk and End buttons are only displayed when the phone is in call mode. And since there are no dedicated Talk and End buttons, you must use a few taps to find these features. That also means you cannot just start dialing a number; you must open the dialpad first, which adds clicks to the process. The same goes for the music player: since there are no external buttons, you must call up the player interface to control your tunes. For some people, the switching back and forth may be a nonissue. But for mutlitaskers, it can grow wearisome.

Still, the interface and keyboard have a long way to go to achieve greatness. For starters, when typing an e-mail or text message the keyboard is displayed only when you hold the iPhone vertically. As a result,

Please select some text.

If the current selection contains information that is also in the known sources, you may mark the text as such using "Mark as known" button.

If you're happy with the current selection, you can add it to one an existing nuggets below, or you may create a new nugget.

Known sources

Assessing selected snippet

Mark as known

Add to new nugget

only 2MP camera and no flash	1	Add
limited to AT T subscription	1	Add
no speed dialing	2	Add
songs can't be used as ringtones	1	Add
requires visual attention; no tactile control	1	Add

Assessment (3)

Id	Topic title	Assessor	Languages	# known sources	total # snippets	# known spans	# marked spans	# chars in marked spans
1	Big Bang Theory	A	EN	3	258	2	164	36591
2	Symptoms Avian Influenza or bird flu	A	EN	2	384	0	46	12595
3	The architecture of Borobudur Temple	A	EN	2	249	0	29	12198
4	Sistemas de calefacción por Biomasa	B	ES,EN,PT,IT	6	324	2	5	3739
5	Sistemas biométricos de autentificación	B	ES,EN,PT,IT	6	241	7	17	4714
6	revistas cientficas open access	B	ES,EN,PT,IT	5	341	4	13	1586
7	Magnum Opus	C	EN	1	308	3	3	765
8	Bloomsday (Band)	C	EN	1	261	6	4	596
9	Belayneh Densamo	C	EN	1	412	16	1	197
10	The Empire in Africa	C	EN	2	235	3	25	6402
11	Gaelic Games	C	EN	1	261	17	11	2706
12	schatten van voorwaardelijke kansen	D	NL	1	291	14	0	0
13	sentiment analysis for European languages other than English	D	NL,EN	1	254	4	2	450
14	European blog search engines	D	NL,EN	1	273	0	6	497
15	verhuistips	D	NL	2	238	26	4	948
16	Yasujiro Ozu	D	NL,EN	3	268	10	10	4570
17	Visa regulations regarding HIV status of applicants	E	EN,NL	0	269	0	31	6237
18	Holidays on Maldives	E	EN	0	281	0	21	1798
19	Comparison of retrieval models in Information Retrieval	E	EN	2	238	0	29	5915
20	ATM (automated teller machine) fraud	E	EN,NL	1	264	0	72	13158
21	iPhone opinions	E	EN,NL	0	290	0	35	4388
22	school education in The Netherlands	E	EN,NL	1	251	9	16	4476
23	Details on obtaining Russian turist visa for foreigners	E	EN,NL	0	285	0	39	7553
24	Albrecht Drer's "Passions" engravings and woodcuts	E	EN	1	387	0	6	807
25	Human-induced climate change: pro and cons	E	EN,NL	0	260	0	27	7345
26	Plastic tableware and the environment	E	EN,NL	0	275	0	26	3887
27	Details on interpretation of Maurice Ravel's "Gaspard de la Nuit"	E	EN	1	250	4	21	8024
28	Nabokov's "Invitation to a Beheading"	E	EN	1	258	11	20	2702
29	Durability of digital storage media	E	EN	0	279	0	9	1605
30	Madonna's books for children	E	EN	1	253	0	45	9019

Assessment (3)

- S
  C

Id	Topic title	Assessor	Languages	# known sources	total # snippets	# known spans	# marked spans	# chars in marked spans
1	Big Bang Theory	A	EN	3	258	2	164	36591
2	Symptoms Avian Influenza or bird flu	A	EN	2	384	0	46	12595
3	The architecture of Borobudur Temple	A	EN	2	249	0	29	12198
4	Sistemas de calefacción por Biomasa	B	ES,EN,PT,IT	6	324	2	5	3739
5	Sistemas biométricos de autentificación	B	ES,EN,PT,IT	6	241	7	17	4714
6	revistas cientficas open access	B	ES,EN,PT,IT	5	341	4	13	1586
7	Magnum Opus	C	EN	1	308	3	3	765
8	Bloomsday (Band)	C	EN	1	261	6	4	596
9	Belayneh Densamo	C	EN	1	412	16	1	197
10	The Empire in Africa	C	EN	2	235	3	25	6402
11	Gaelic Games	C	EN	1	261	17	11	2706
12	schatten van voorwaardelijke kansen	D	NL	1	291	14	0	0
13	sentiment analysis for European languages other than English	D	NL,EN	1	254	4	2	450
14	European blog search engines	D	NL,EN	1	273	0	6	497
15	verhuistips	D	NL	2	238	26	4	948
16	Yasujiro Ozu	D	NL,EN	3	268	10	10	4570
17	Visa regulations regarding HIV status of applicants	E	EN,NL	0	269	0	31	6237
18	Holidays on Maldives	E	EN	0	281	0	21	1798
19	Comparison of retrieval models in Information Retrieval	E	EN	2	238	0	29	5915
20	ATM (automated teller machine) fraud	E	EN,NL	1	264	0	72	13158
21	iPhone opinions	E	EN,NL	0	290	0	35	4388
22	school education in The Netherlands	E	EN,NL	1	251	9	16	4476
23	Details on obtaining Russian turist visa for foreigners	E	EN,NL	0	285	0	39	7553
24	Albrecht Drer's "Passions" engravings and woodcuts	E	EN	1	387	0	6	807
25	Human-induced climate change: pro and cons	E	EN,NL	0	260	0	27	7345
26	Plastic tableware and the environment	E	EN,NL	0	275	0	26	3887
27	Details on interpretation of Maurice Ravel's "Gaspard de la Nuit"	E	EN	1	250	4	21	8024
28	Nabokov's "Invitation to a Beheading"	E	EN	1	258	11	20	2702
29	Durability of digital storage media	E	EN	0	279	0	9	1605
30	Madonna's books for children	E	EN	1	253	0	45	9019

Evaluation measures

Evaluation measures

- Based on standard precision and recall

Evaluation measures

- Based on standard precision and recall
- For a given response *R* (ranked list of snippets) of a system *S* for topic *T*, define
  - **recall** is the sum of character lengths of all spans in *R* linked to nuggets, divided by total sum of span lengths
  - **precision** is the number of characters that belong to at least one span linked to a nugget, divided by total character length of *R*

Runs

Runs

- What did people try?
  - DCU
    - Sentence-based snippets. Multiple ways of re-ranking of snippets: (1) word overlap with topic, description, known source; (2) word overlap plus thresholding; (3) compare parses of known sources with parses of snippets.
  - UIndonesia
    - …
  - USAL
    - Fixed size text windows (1500 bytes). Focus on segmentation (“snippet generation”); ranking based on structured queries (topic, description, anchor text, the vocabulary from the “known sources”)
  - UvA
    - Sentence-based and paragraph-based snippets. Centrality scores plus penalties for overlap with known sources (see next talk).

Runs (2)

Participant	Run	Average snippet length	Average snippets per topic	Average response length per topic
Baseline	Google snippets	145	898	131041
School of Computing, Dublin City University	DCU run1 simple	118	30	3552
	DCU run2 parsed	137	27	3770
	DCU run2 topfilter	112	29	3346
Faculty of Computer Science, University of Indonesia	UIWC07odwgstr	151	10	1522
	UIWC07uw10	155	10	1559
	UIWC07wstr	152	10	1530
REINA Research Group, University of Salamanca	USAL reina0.25	833	50	41680
	USAL reina0	832	50	41658
	USAL reina1	833	50	41708
ISLA, University of Amsterdam	UVA par vs	254	29	7420
	UVA par wo	277	25	7158
	UVA sent wo	214	33	7225

Runs (2)

Participant	Run	Average snippet length	Average snippets per topic	Average response length per topic
Baseline	Google snippets	145	898	131041
School of Computing, Dublin City University	DCU run1 simple	118	30	3552
	DCU run2 parsed	137	27	3770
	DCU run2 topfilter	112	29	3346
Faculty of Computer Science, University of Indonesia	UIWC07odwgstr	151	10	1522
	UIWC07uw10	155	10	1559
	UIWC07wstr	152	10	1530
REINA Research Group, University of Salamanca	USAL reina0.25	833	50	41680
	USAL reina0	832	50	41658
	USAL reina1	833	50	41708
ISLA, University of Amsterdam	UVA par vs	254	29	7420
	UVA par wo	277	25	7158
	UVA sent wo	214	33	7225

Runs (2)

Participant	Run	Average snippet length	Average snippets per topic	Average response length per topic
Baseline	Google snippets	145	898	131041
School of Computing, Dublin City University	DCU run1 simple	118	30	3552
	DCU run2 parsed	137	27	3770
	DCU run2 topfilter	112	29	3346
Faculty of Computer Science, University of Indonesia	UIWC07odwgstr	151	10	1522
	UIWC07uw10	155	10	1559
	UIWC07wstr	152	10	1530
REINA Research Group, University of Salamanca	USAL reina0.25	833	50	41680
	USAL reina0	832	50	41658
	USAL reina1	833	50	41708
ISLA, University of Amsterdam	UVA par vs	254	29	7420
	UVA par wo	277	25	7158
	UVA sent wo	214	33	7225

Runs (2)

Participant	Run	Average snippet length	Average snippets per topic	Average response length per topic
Baseline	Google snippets	145	898	131041
School of Computing, Dublin City University	DCU run1 simple	118	30	3552
	DCU run2 parsed	137	27	3770
	DCU run2 topfilter	112	29	3346
Faculty of Computer Science, University of Indonesia	UIWC07odwgstr	151	10	1522
	UIWC07uw10	155	10	1559
	UIWC07wstr	152	10	1530
REINA Research Group, University of Salamanca	USAL reina0.25	833	50	41680
	USAL reina0	832	50	41658
	USAL reina1	833	50	41708
ISLA, University of Amsterdam	UVA par vs	254	29	7420
	UVA par wo	277	25	7158
	UVA sent wo	214	33	7225

Results

Run	@ 1,500 bytes		@ 3,500 bytes		@ 7,000 bytes	
	P	R	P	R	P	R
Google snippets	0.13	0.3	0.11	0.07	0.08	0.11
DCU run1 simple	0.07	0.02	0.08	0.05	–	–
DCU run2 parsed	0.10	0.03	0.10	0.06	–	–
DCU run2 topfilter	0.08	0.02	0.08	0.04	–	–
UIWC07odwgstr	0.11	0.03	–	–	–	–
UIWC07uw10	0.09	0.02	–	–	–	–
UIWC07wstr	0.11	0.03	–	–	–	–
USAL reina0.25	0.11	0.03	0.14	0.09	0.16	0.20
USAL reina0	0.11	0.03	0.13	0.08	0.14	0.18
USAL reina1	0.11	0.03	0.14	0.09	0.16	0.21
UVA par vs	0.19	0.05	0.20	0.13	0.20	0.26
UVA par wo	0.15	0.04	0.20	0.13	0.20	0.25
UVA sent wo	0.10	0.03	0.09	0.06	0.09	0.11

Results

- Baseline plus P/R values at three cut-off points

Run	@ 1,500 bytes		@ 3,500 bytes		@ 7,000 bytes	
	P	R	P	R	P	R
Google snippets	0.13	0.3	0.11	0.07	0.08	0.11
DCU run1 simple	0.07	0.02	0.08	0.05	–	–
DCU run2 parsed	0.10	0.03	0.10	0.06	–	–
DCU run2 topfilter	0.08	0.02	0.08	0.04	–	–
UIWC07odwgstr	0.11	0.03	–	–	–	–
UIWC07uw10	0.09	0.02	–	–	–	–
UIWC07wstr	0.11	0.03	–	–	–	–
USAL reina0.25	0.11	0.03	0.14	0.09	0.16	0.20
USAL reina0	0.11	0.03	0.13	0.08	0.14	0.18
USAL reina1	0.11	0.03	0.14	0.09	0.16	0.21
UVA par vs	0.19	0.05	0.20	0.13	0.20	0.26
UVA par wo	0.15	0.04	0.20	0.13	0.20	0.25
UVA sent wo	0.10	0.03	0.09	0.06	0.09	0.11

Results (2)

Results (2)

- *UVA par vs* and *UVA par wo* best performance across all cut-off points

Results (2)

- *UVA par vs* and *UVA par wo* best performance across all cut-off points
- *USAL reina0.25* and *USAL reina1* have comparable performance

Results (2)

- *UVA par vs* and *UVA par wo* best performance across all cut-off points
- *USAL reina0.25* and *USAL reina1* have comparable performance

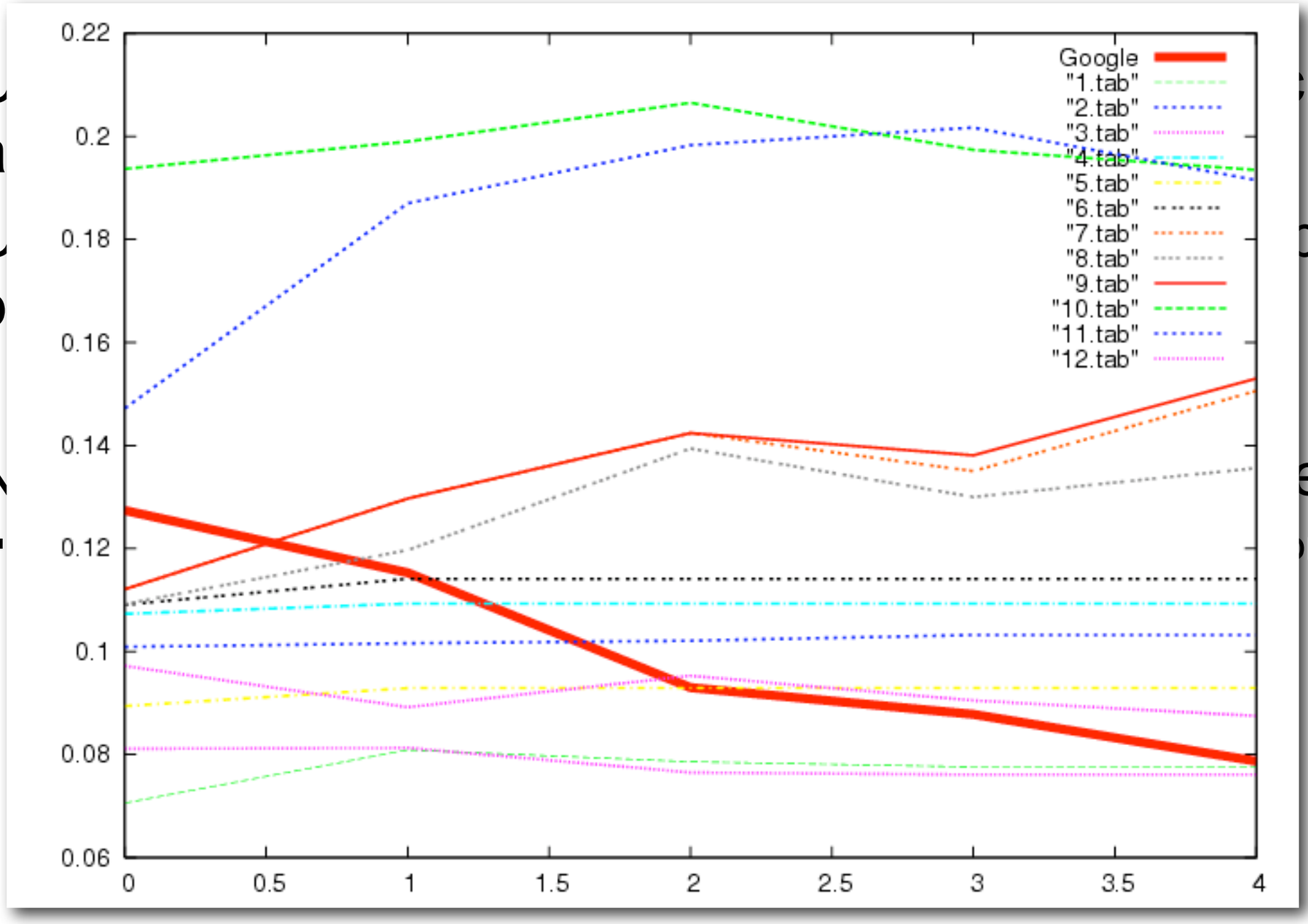
Results (2)

- *UVA par vs* and *UVA par wo* best performance across all cut-off points
- *USAL reina0.25* and *USAL reina1* have comparable performance

- Note: precision grows as the cut-off point increases
  - Systems manage to find relevant snippets, but the ranking is far from optimal

Results (2)

- *U*
  a
- *U*
  p
- N
  -



cross

ble

es

om

Conclusions

Conclusions

- WebCLEF 2007
  - New task
  - Aimed at undirected informational search goals (“Tell me about X.”)
  - Most submitted runs outperformed the Google-based baseline

Conclusions

- WebCLEF 2007
  - New task
  - Aimed at undirected informational search goals (“Tell me about X.”)
  - Most submitted runs outperformed the Google-based baseline

- Work left to be done
  - Detailed error analysis
  - Refine assessor guidelines
  - Refine evaluation interface
  - Not enough topics?

Conclusions

- WebCLEF 2007
  - New task
  - Aimed at undirected informational search goals (“Tell me about X.”)
  - Most submitted runs outperformed the Google-based baseline

- Work left to be done
  - Detailed error analysis
  - Refine assessor guidelines
  - Refine evaluation interface
  - Not enough topics?

- Thanks
  - Participants and assessors

WebCLEF — The Future?

WebCLEF — The Future?

- Why does the most interesting task at CLEF have so few participants?

WebCLEF — The Future?

- Why does the most interesting task at CLEF have so few participants?

WebCLEF — The Future?

- Why does the most interesting task at CLEF have so few participants?
- The track will only continue if the number of participants is sufficiently

WebCLEF — The Future?

- Why does the most interesting task at CLEF have so few participants?
- The track will only continue if the number of participants is sufficiently

WebCLEF — The Future?

- Why does the most interesting task at CLEF have so few participants?
- The track will only continue if the number of participants is sufficiently
- To help you prepare for next year
  - 2007 topic, documents, qrels
  - Code of the best performing 2007 system freely available (as a baseline)

WebCLEF — The Future?

- Why does the most interesting task at CLEF have so few participants?
- The track will only continue if the number of participants is sufficiently
- To help you prepare for next year
  - 2007 topic, documents, qrels
  - Code of the best performing 2007 system freely available (as a baseline)
- More at the breakout session
  - (12:00–13:00, this room)

Tips

Tips

- Look at summarization
  - Evaluation measures
  - Especially last two years

Tips

- Look at summarization
  - Evaluation measures
  - Especially last two years
- Passages vs paragraphs
  -