Università della Svizzera italiana

Facoltà di scienze informatiche

Information Retrieval In Context

Prof. Fabio Crestani
fabio.crestani@usi.ch

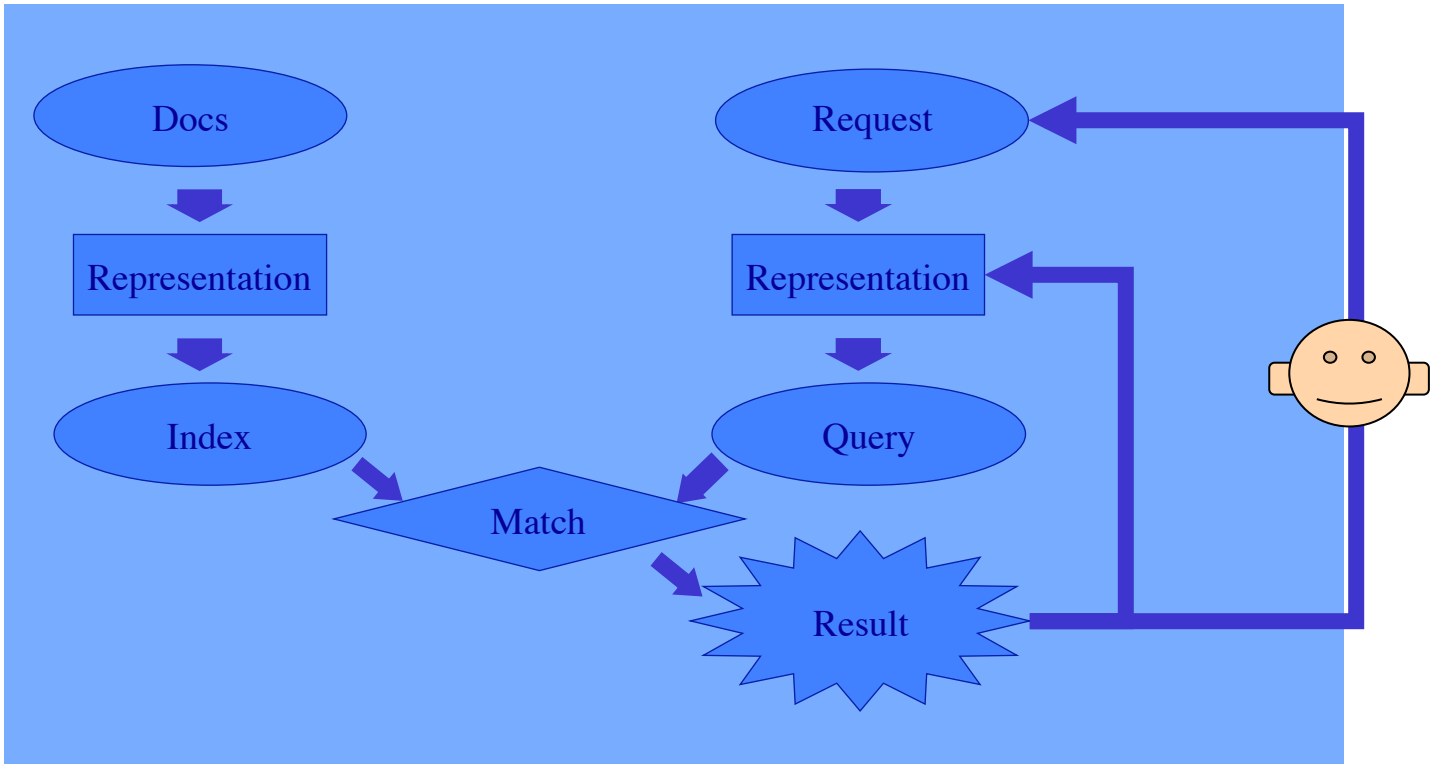
Outline

Information Retrieval (IR)
IR in context (and the meaning of context in IR)
The many roles of context in IR
Example of the use of context in IR

1. Determining the importance of context for mobile IR
2. Search and visualisation of contextual information in hierarchically structured documents
3. Personalised and context-aware document summarisation for mobile IR
4. User perception of relevance of spoken documents for mobile IR

Information Retrieval

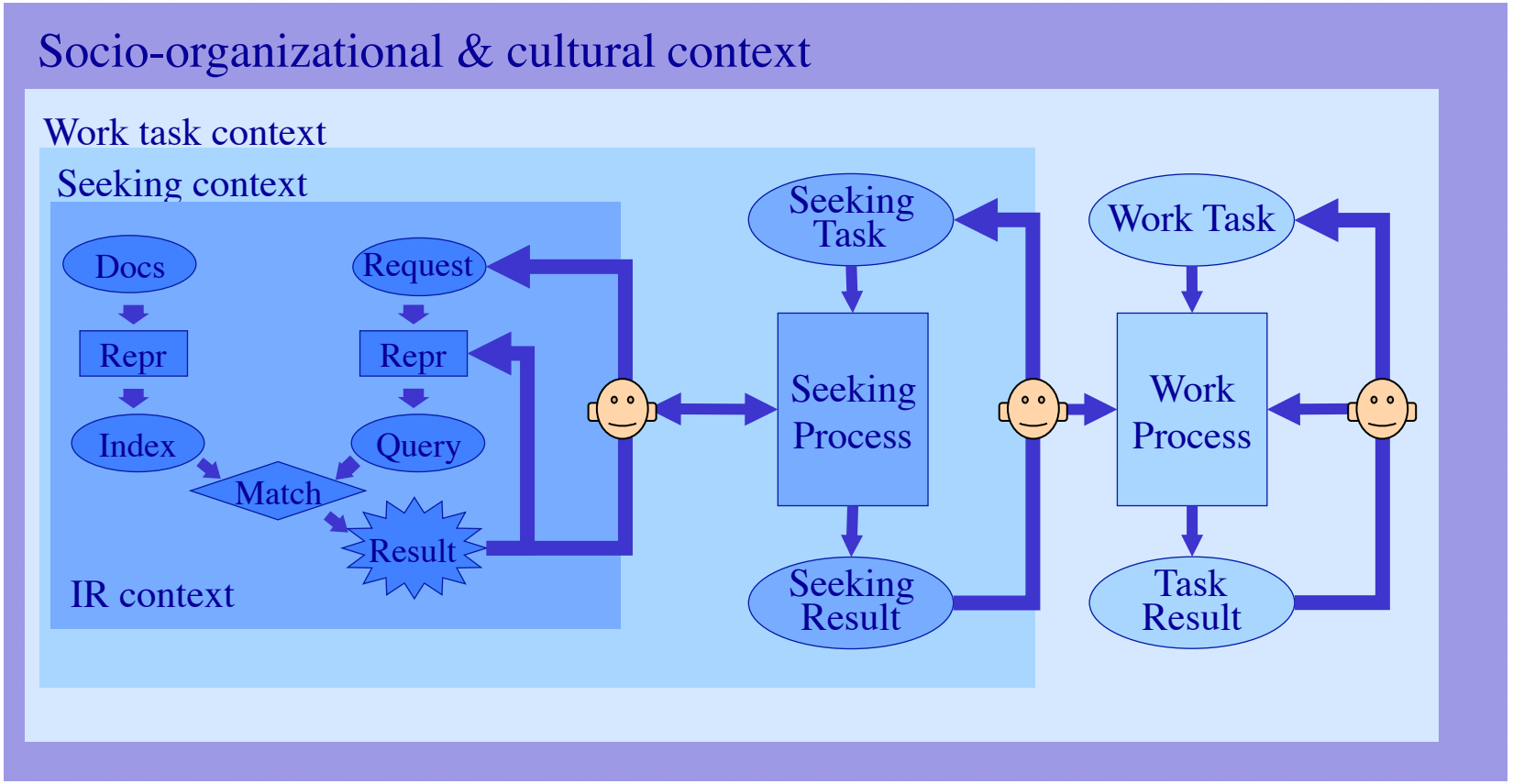
What is IR?



Where is the context?

IR in context

Where is context?



How much context do we need?

Clearly, current IR systems do not take into account all this context

How can we design IR systems that do that? How much context should we consider?

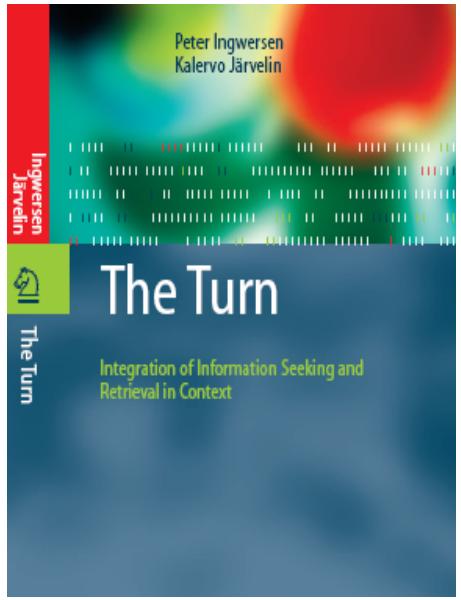
- In principle, we should consider whatever has important relationships with the application of IR
- however, what has, is not well known
- Answering this requires theoretical modelling, implementation and empirical testing (evaluation) of context

Different approaches …

How much context do we need?

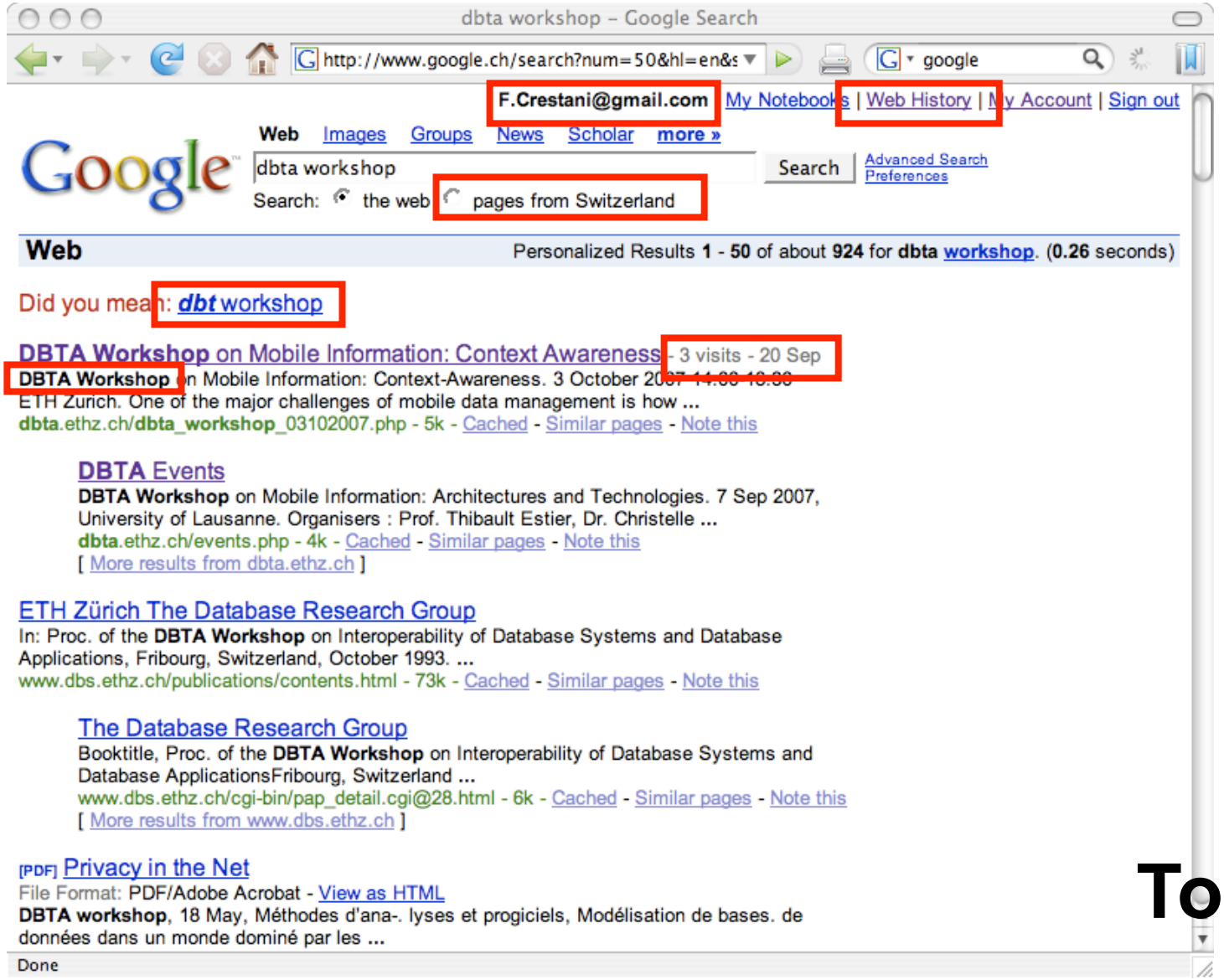
Different dimensions of context:

1. Work task dimension
2. Search task dimension
3. Actor dimension
4. Perceived work task dimension
5. Perceived search task
6. Document dimension
7. Algorithmic search engine dimension
8. Algorithmic interface dimension
9. Access and interaction dimension



Too much?

How much context do we need?



Too little?

How much context?

How much context do you use?

What elements of context provided by Google in the result presentation do you actually use?

Can you identify specific types of queries for which these contextual information might be useful?

Where do we need context?

We can use context at different stages of the IR process:

- Query specification
- Retrieval (matching/ranking)
- Results presentation and interaction
- …

Context should shape these stages by adapting them to the:

- User preferences
- Search task
- Work task
- …

Corollary: IR in Context ≠Personalised IR

- In fact: IR in Context >> Personalised IR

How do we use context?

Assume:

- IR process = Query, Retrieval, Presentation , …,
- IR in context = IR process + Context
- IR in context = (Query, Retrieval, Presentation, …) + Context

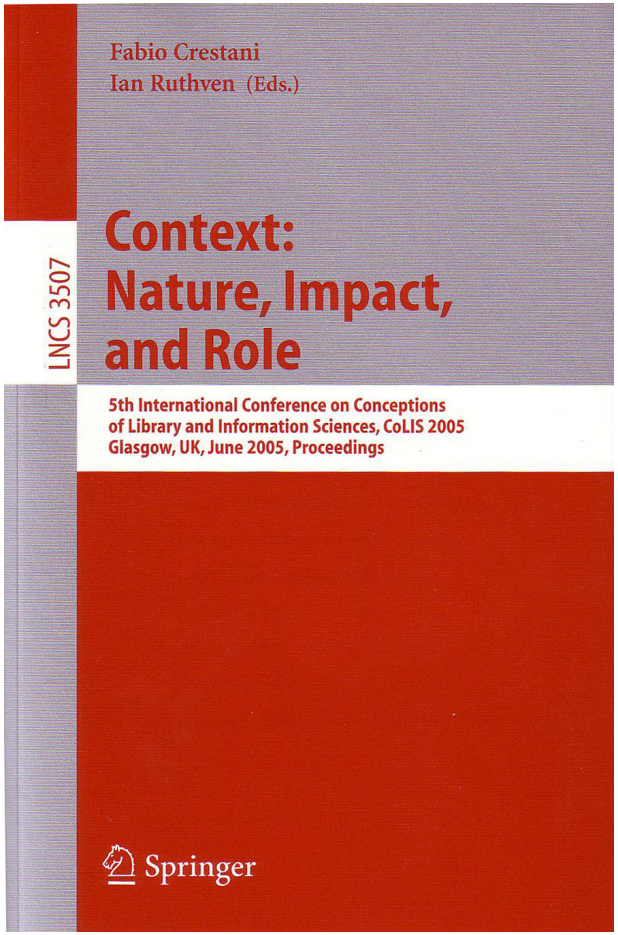
But this is too difficult!

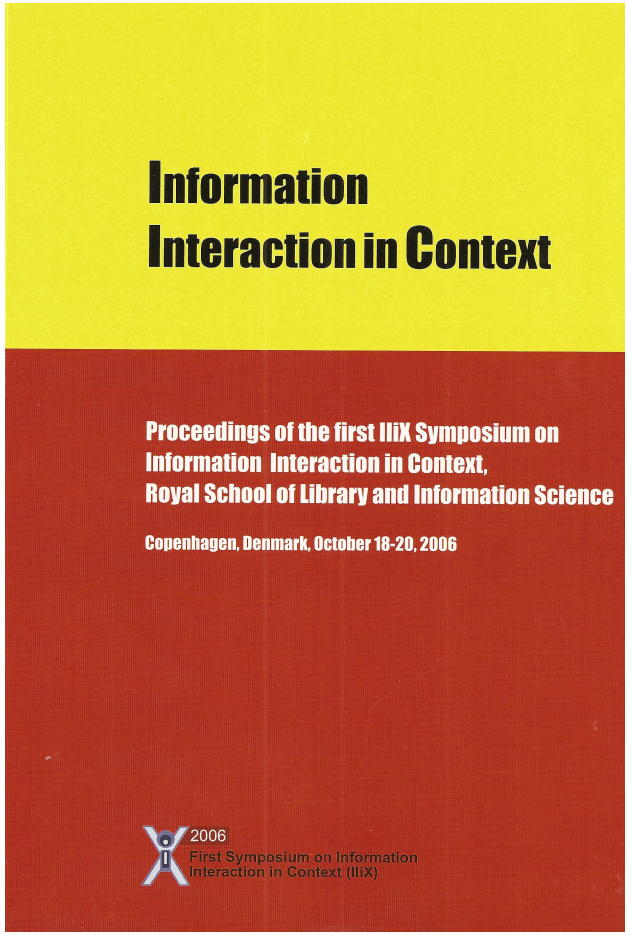
Better start with:

1. IR= (Query + Context), Retrieval, Presentation, …
2. IR= Query, (Retrieval + Context), Presentation, …
3. IR= Query, Retrieval, (Presentation + Context), …

In the following a will present a few examples, mostly of my own work, on *the use of context in 1 & 3*

Commercial break





Context as query modifier

IR= (Query + Context), Retrieval, Presentation, …

Thus context is additional information that should/ could enhance the query

But what information?

Context for Mobile IR

Consider web IR, what additional information provided by the context could modify the query?

Let us now consider mobile IR, what features of context could modify the query?

Dimensions of mob. IR context

There is some consensus in the literature on the following elements (or dimensions) of context for Mobile IR:

1. Profile
2. Location
3. Time
4. Activity
5. Agenda
6. Service
7. Preferences
8. Situation
9. Environment
10. Social Context

Context for Mobile IR

We carried out a *crowdsourcing* experiment to see what users thought were the most important contextual dimensions for mobile IR

We user over 60 queries of different types:

- Mobile IR vs. TREC vs. web
- Informational vs. transactional vs. navigational
- 10 dimensions of context

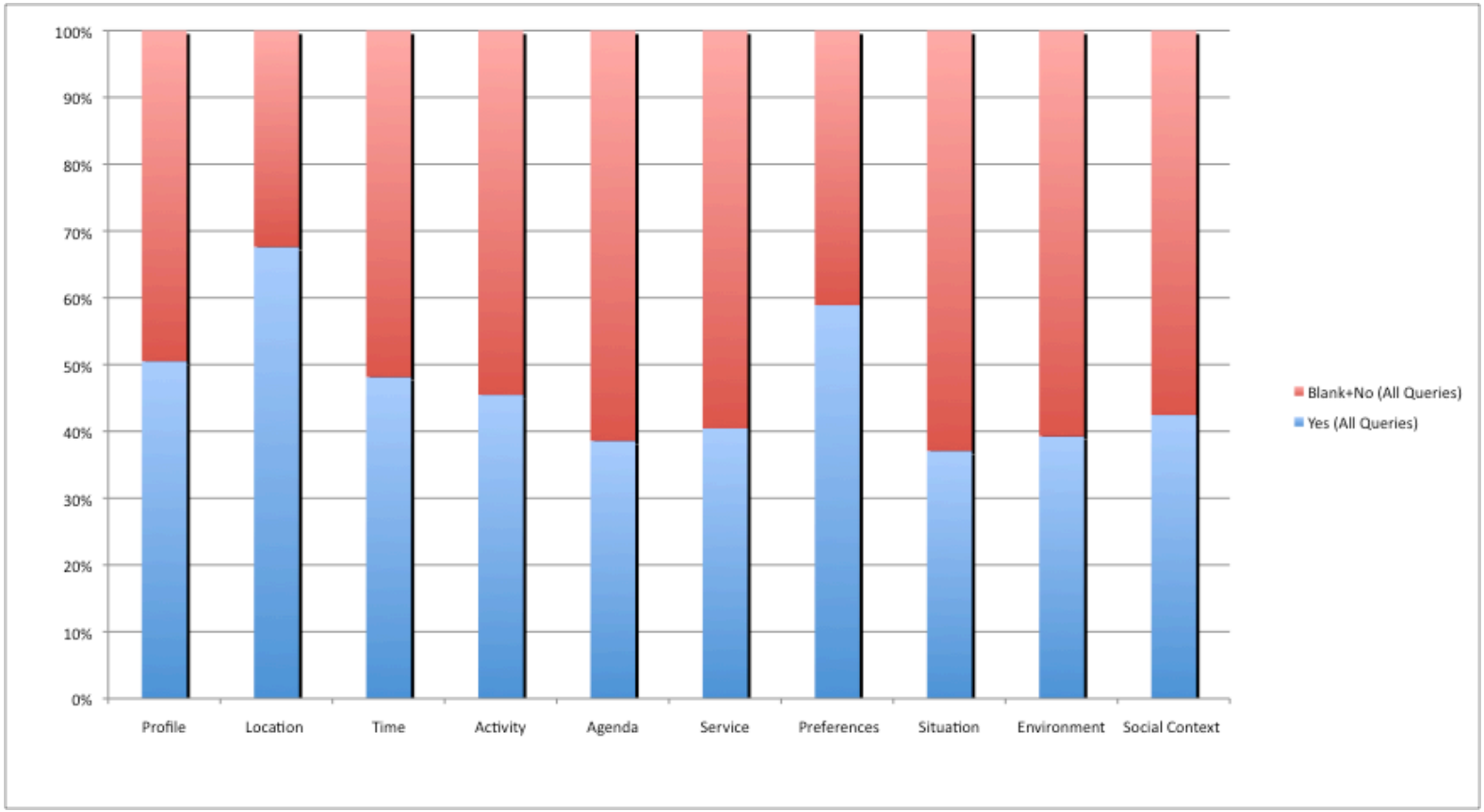
The question to users was:

- "Is this element of context useful to answer this query? Why?

The experiment involved over 200 users with at least 20 judgements for each query

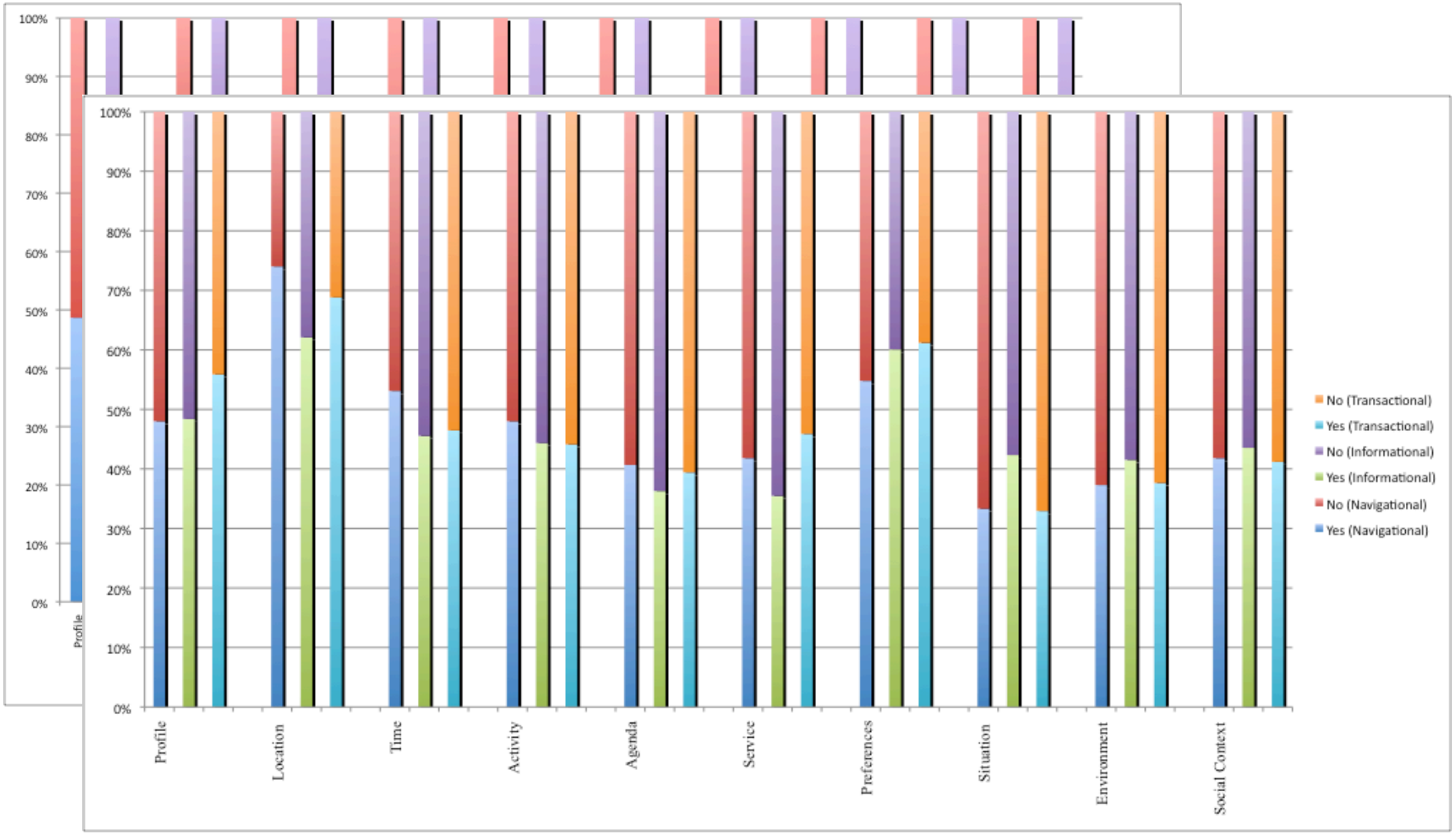
Importance of context in mob. IR

What are the most important elements?



Importance of context in mob. IR

Does it depend on the type of query?



Conclusions on context for mob. IR

Context is certainly important …

… but it is not clear what elements of context are useful, when and why!

Results presentation in context

Results presentation is a complex stage of the IR process:

- The query is often an imprecise expression of the users information need
- Results are often presented using document surrogates
- Correct relevance assessment is needed to trigger query reformulation and relevance feedback
- The assessment of the quality of the results has the highest impact on the perceived quality of the IR system

Results presentation in context can make relevance assessment more effective

Context for results presentation

In results presentation context can have different roles, like for example:

- As “additional information” to be provided to the user to help assess the relevance of the results presented (especially when results are complex information objects)
- As a “modifier of the results” that are adapted to the context (search task, work task, device, user preferences, …)

I will briefly present some examples of each of these uses of context in results presentation

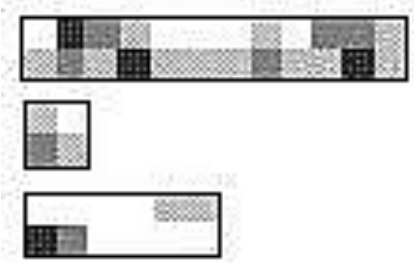
Context as additional information

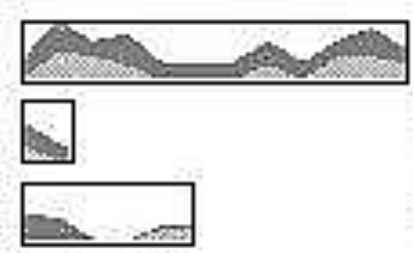
Simple and intuitive definition :

- Context is any additional information that enhances the understanding of the document being presented as result of a search
- This includes *data*, e.g. other information in the same document or in other documents, and *metadata*, e.g. the position of the information in the document, the relation with other sections of the document, the author and title of the section or the document, etc.

Visualising context

Many different approaches proposed in the past in the document visualisation research area
Some approaches are directed to specific types of documents and specific applications
Best general approaches for IR: Tilebars, Relevance Curves and Thumbnails





Tilebars and relevance curves

Tilebar

Relevance Curve

Show the relevance of each passage or segment of text

- Shows the document size
- Does not show the hierarchical relations of the document structural elements
- Does not show the context of the relevant elements

Thumbnails

Thumbnail

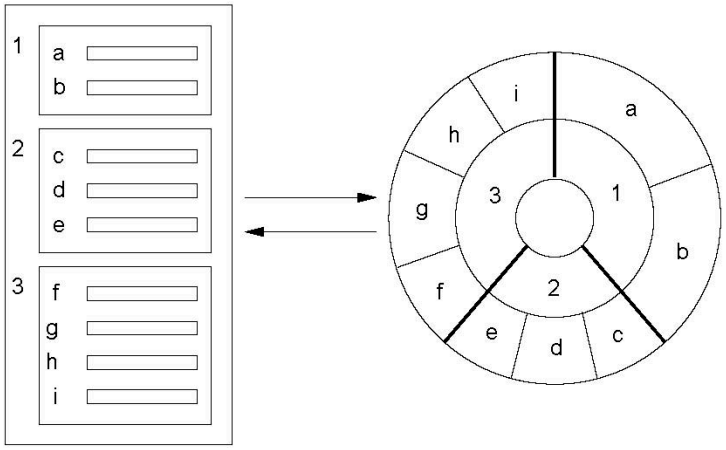
Show the appearance of the document:

- Useful when the user works frequently with the same document set and can recognise a document by its thumbnail view.
- Could show context (e.g. KWIC)
- Does not show the size of the document
- Does not show the structure of the document
- Does not show the relevance of each element of the document
- Could be a good complement to Tilebars and Relevance Curves

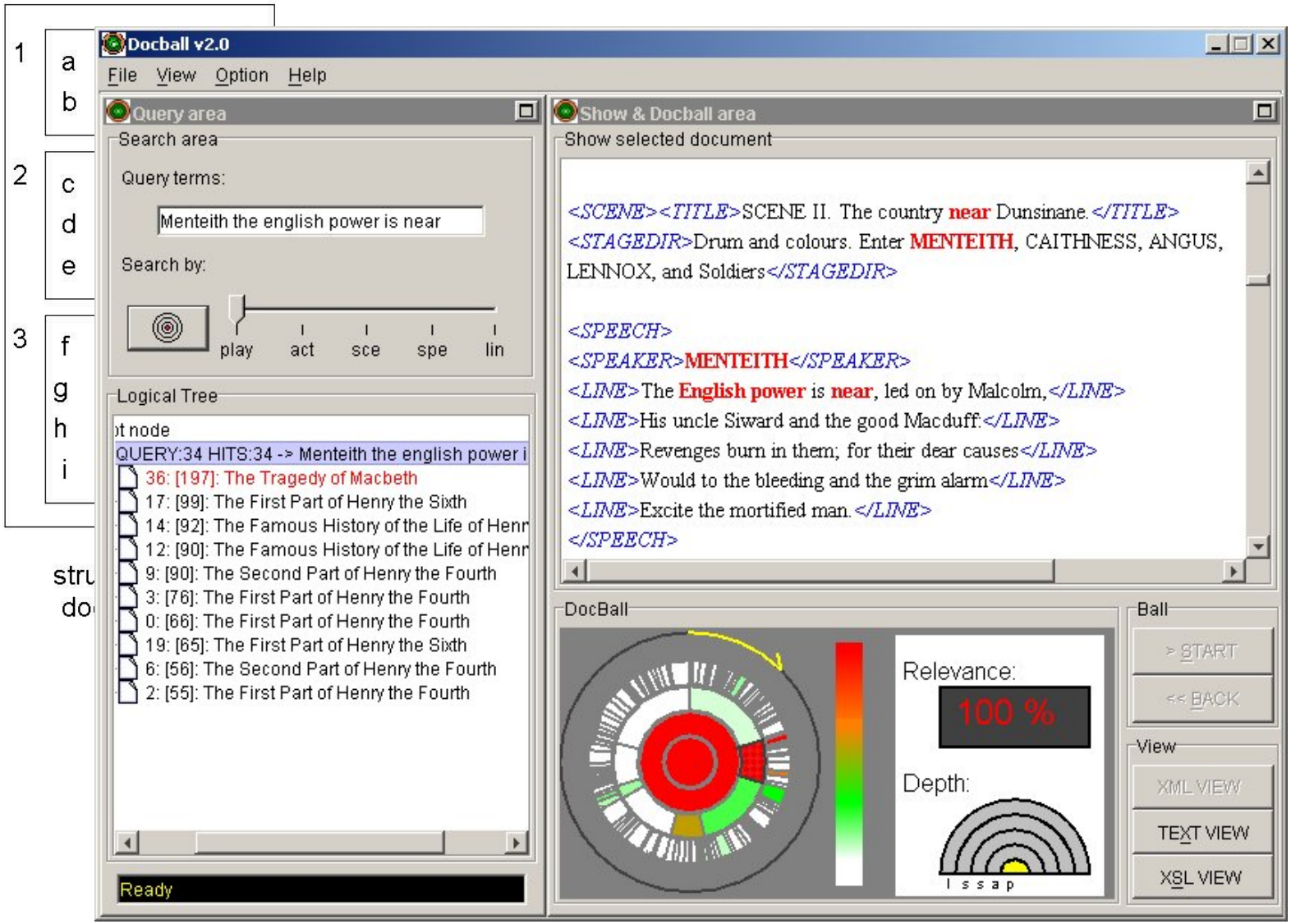
The Docball metaphor

We developed a visualisation metaphor that enables to show the structure, size, relevance of each element in the structure, and context

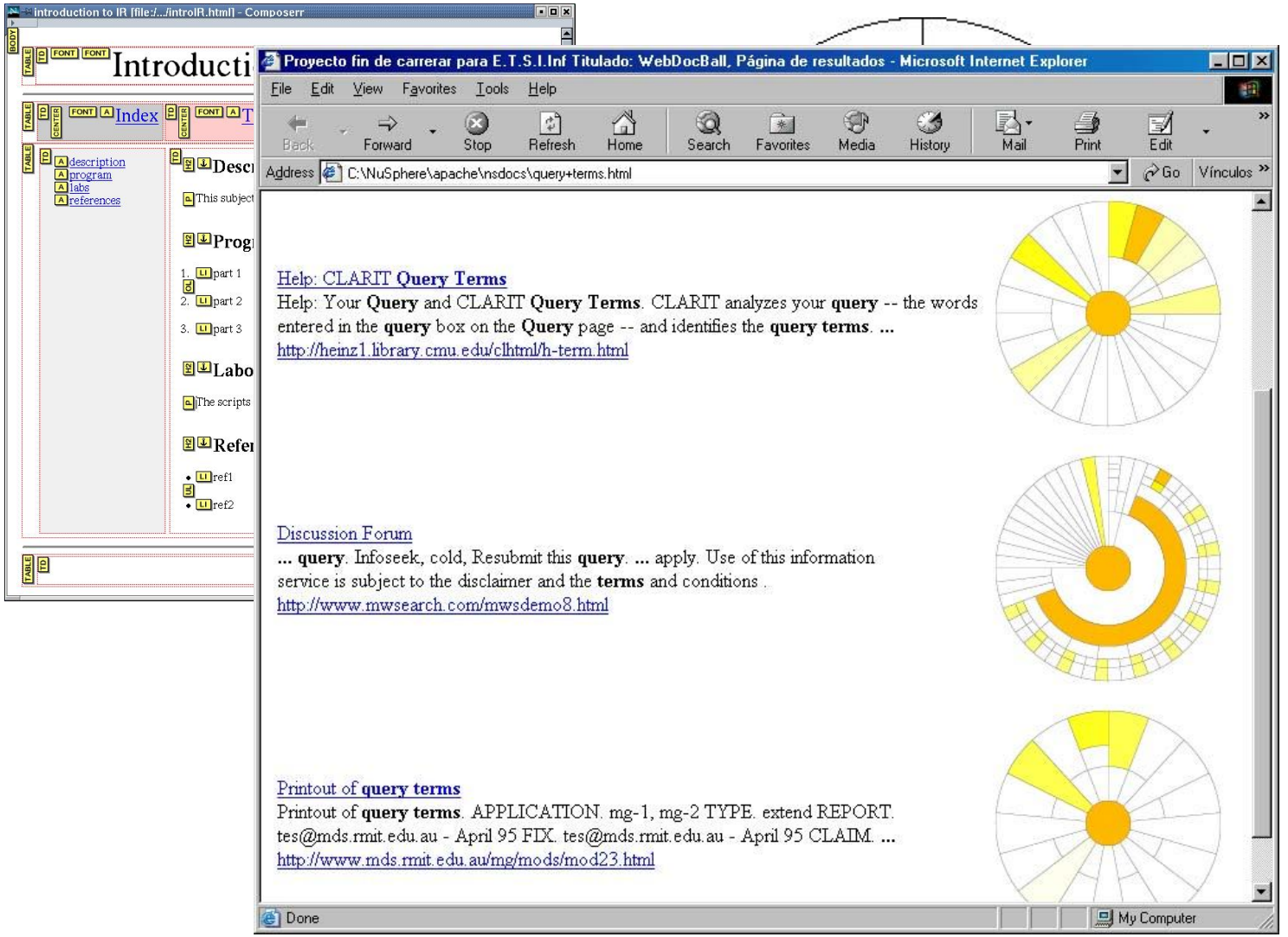
Simple idea:



DocBall


1
a
b
2
c
d
e
3
f
g
h
i
Docball v2.0
File View Option Help
Query area
Search area
Query terms:
Menteith the english power is near
Search by:
play act sce spe lin
Logical Tree
QUERY:34 HITS:34 -> Menteith the english power i
36: [197]: The Tragedy of Macbeth
17: [99]: The First Part of Henry the Sixth
14: [92]: The Famous History of the Life of Henr
12: [90]: The Famous History of the Life of Henr
9: [90]: The Second Part of Henry the Fourth
3: [76]: The First Part of Henry the Fourth
0: [66]: The First Part of Henry the Fourth
19: [65]: The First Part of Henry the Sixth
6: [56]: The Second Part of Henry the Fourth
2: [55]: The First Part of Henry the Fourth
Ready
Show & Docball area
Show selected document
<SCENE><TITLE>SCENE II. The country near Dunsinane.</TITLE>
<STAGEDIR>Drum and colours. Enter MENTEITH, CAITHNESS, ANGUS, LENNOX, and Soldiers</STAGEDIR>
<SPEECH>
<SPEAKER>MENTEITH</SPEAKER>
<LINE>The English power is near, led on by Malcolm,</LINE>
<LINE>His uncle Siward and the good Macduff:</LINE>
<LINE>Revenges burn in them; for their dear causes</LINE>
<LINE>Would to the bleeding and the grim alarm</LINE>
<LINE>Excite the mortified man.</LINE>
</SPEECH>
DocBall
Relevance:
100 %
Depth:
Issap
Ball
START
BACK
View
XML VIEW
TEXT VIEW
XSL VIEW

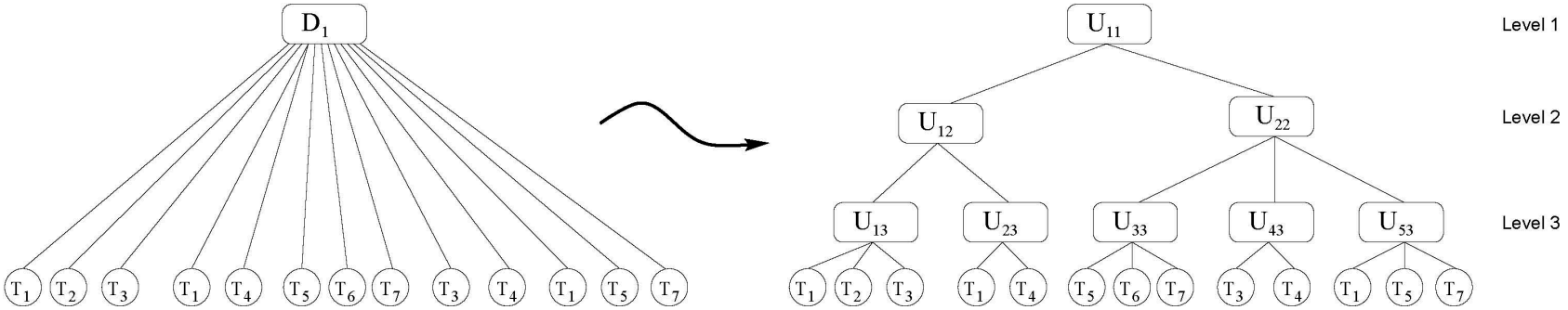
WebDocBall


introduction to IR [file:/.../introIR.html] - Composerr
Introducti
Index
description
program
labs
references
Proyecto fin de carrerar para E.T.S.I.Inf Titulado: WebDocBall, Página de resultados - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Back Forward Stop Refresh Home Search Favorites Media History Mail Print Edit
Address C:\NuSphere\apache\nsdocs\query+terms.html
Go Vínculos
Help: CLARIT Query Terms
Help: Your Query and CLARIT Query Terms. CLARIT analyzes your query -- the words entered in the query box on the Query page -- and identifies the query terms. ...
http://heinz1.library.cmu.edu/clhtml/h-term.html
Discussion Forum
... query. Infoseek, cold, Resubmit this query. ... apply. Use of this information service is subject to the disclaimer and the terms and conditions .
http://www.mwsearch.com/mwsdemo8.html
Printout of query terms
Printout of query terms. APPLICATION. mg-1, mg-2 TYPE. extend REPORT. tes@mds.rmit.edu.au - April 95 FIX. tes@mds.rmit.edu.au - April 95 CLAIM. ...
http://www.mds.rmit.edu.au/mg/mods/mod23.html
Done
My Computer

Hierarchically structured docs

Many types of documents are naturally hierarchically structured (HSD)

- Textbooks, manuals, scientific articles, Web pages, …



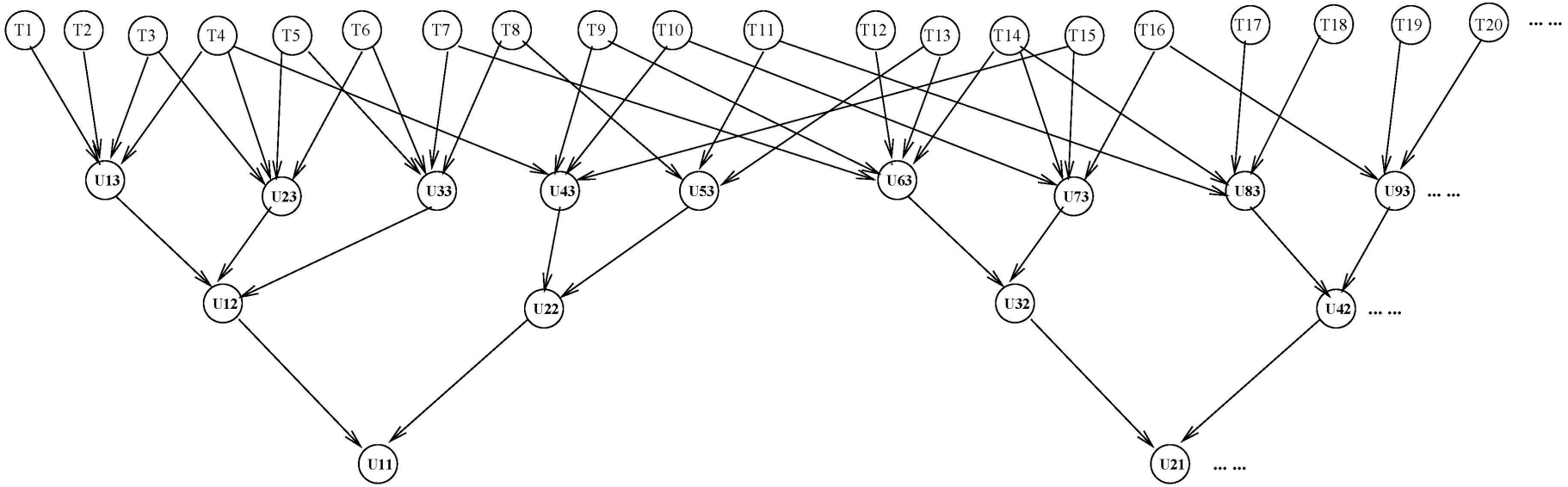
Users are often interested only in some small relevant parts of HSD

Some parts of documents, though not relevant, provide “contextual information” that facilitate the user’s relevance assessment and increases the usefulness of relevant information

Context: where is it?

We developed the *SRIDE^{RB} model*, combining:

- A Multilayered Bayesian Network model -> estimates relevance at each HSD element
- A Utility Theory model -> use Decision theory to decide the best HSD element to present to the user, based on some utility function



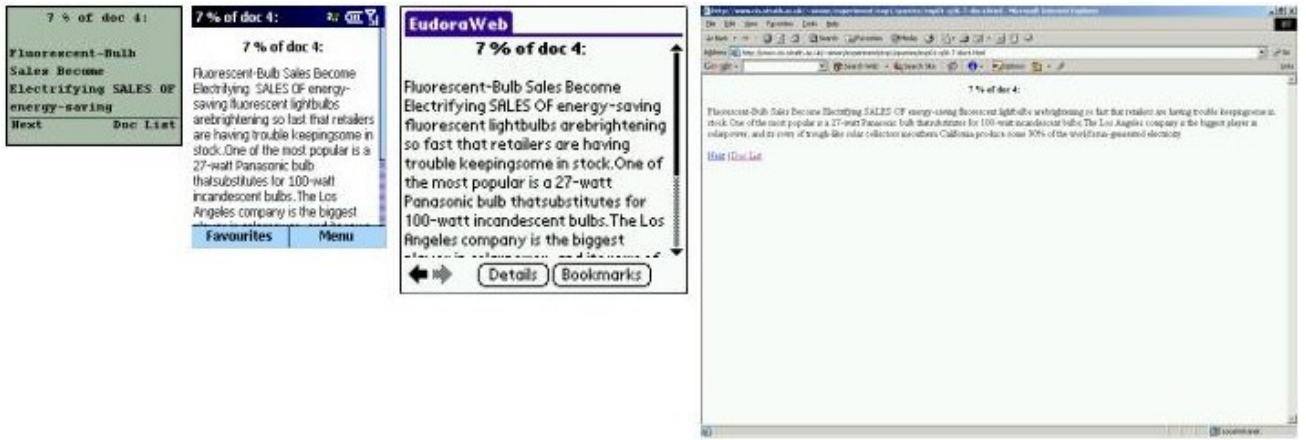
Context as results presentation modifier

The user context determines which results are useful at that particular time
Thus, results presentation should be modified or *adapted to the user context* to help the user assess the relevance of the search results presented

Results pres. for Mobile IR

Mobile IR is the perfect environment to use context as a modifier

- Mobile IR is: personal, location-dependent, time-dependent, device-dependent, …
- Relevance assessment in Mobile-IR is more complex than in standard IR
  - Screen size and information presented



  - Spoken presentation of results …

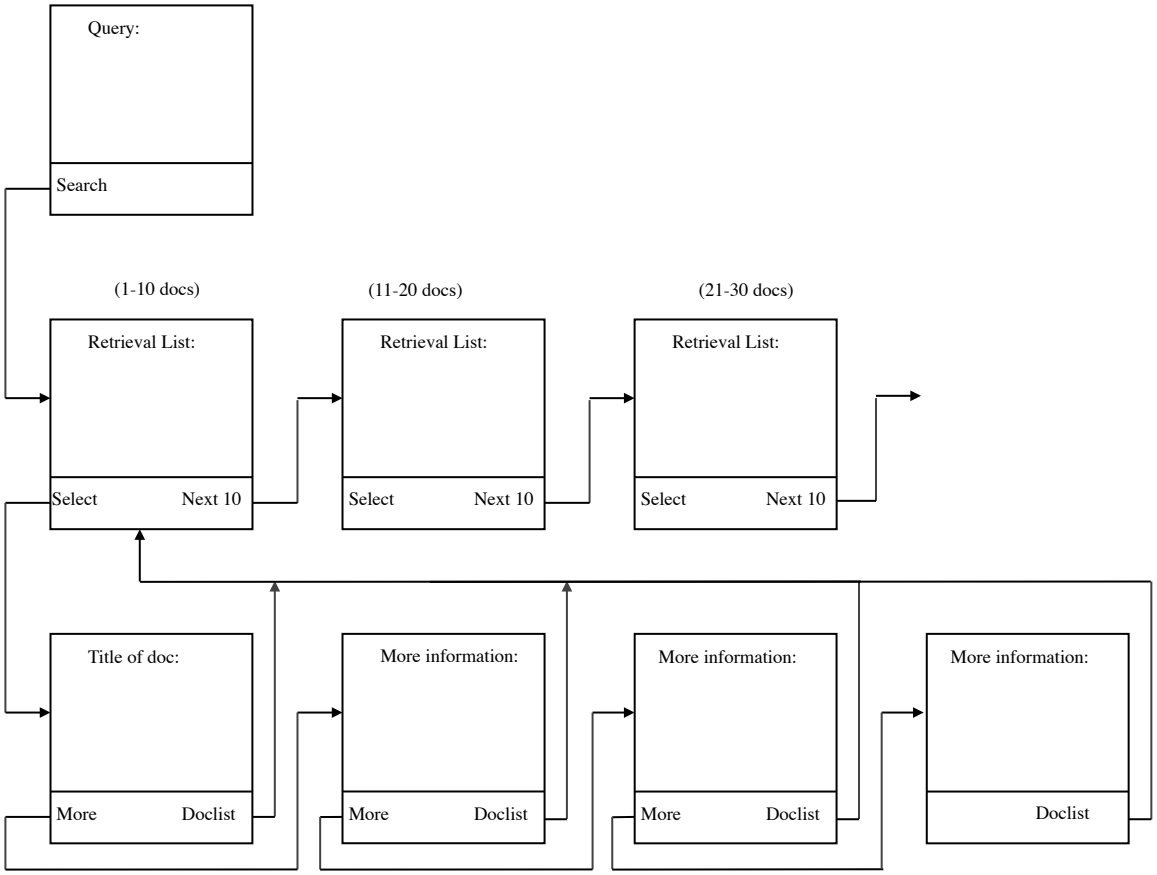
Query-biased Summarisation

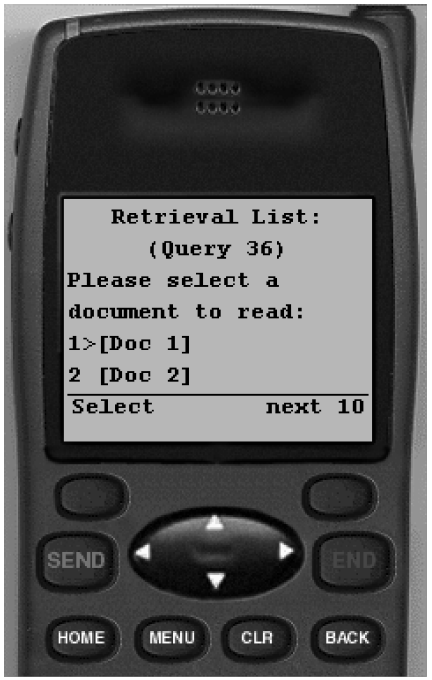
Best technique of automatic text summarisation by sentence extraction for retrieval results presentation (better than full text!)
Sentences are weighted using a combination of evidence from:

- heuristics (e.g.~news title and headings)
- term weights based on collection and document statistics (tf-idf)
- query score (consider terms in query)

Hierarchical summarisation

We developed a new type of summarisation for Mobile IR: *query-biased hierarchical summarisation*





But what is more information?

More information

"More information" is determined by the context

- The query (query-biased summarisation)
- Document type
- Device screen size
- User's media preferences
- Location
- Time
- Work task
- …

We can say that "more information is: *new/novel information* (i.e. not already seen by user)

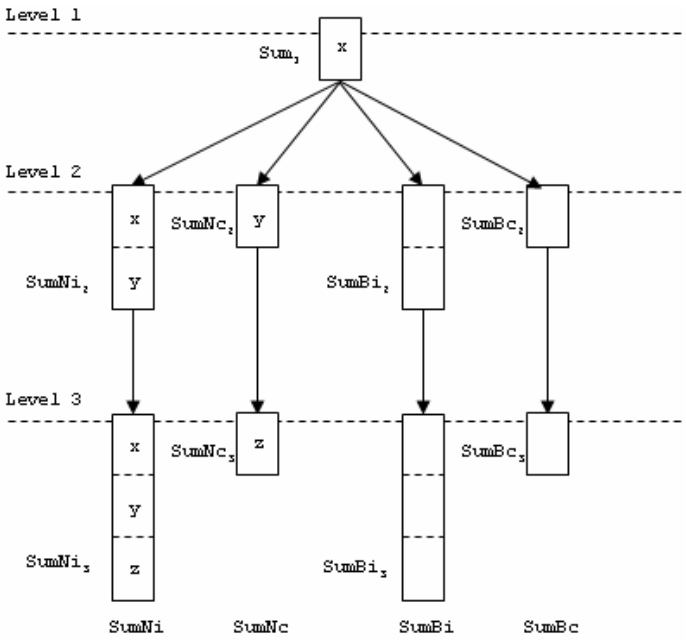
- Should we show this new information in context?

Summarisation in context

We introduced novelty detection in summarisation to determine the new information to show to the user at each “give me more” request

We showed users the novel information together with the information they already received or on its own

- Is there any difference?
- Is there an additional cognitive load?



Yes, but a very small one

Results presentation using speech

Some media are naturally more context rich than others (think of an image or a video wrt text)

Speech contains a lot more information than text

Is this a good thing with regards to the presentation of results?

- Would speech be more effective for results presentation than text?

User's Perception of Relevance

The absence of a screen compel Mobile IR application to present retrieval results in some non traditional way

Presenting retrieval results using spoken document surrogates seems an obvious choice

But:

- Do users make correct relevance assessments when presented with document surrogates?
- Do users make correct relevance assessments when presented with spoken document surrogates?
- Do users make faster or slower assessments?

Perception of Relevance & Speech

We decided to experiment the effectiveness of retrieval results presentation using spoken surrogates

In a previous study we have shown that document summaries are more effective than other surrogates to present retrieval results

- Study carried out on PCs, PDAs and WAP mobile phones
- Study involved different summary lengths and different sentence extraction based summarization techniques

We now want to study spoken summaries

Experimental Settings

Subjects: 10 native English speakers

Text collection: summaries of documents results of 50 topics from TREC collection

Experimental procedure:

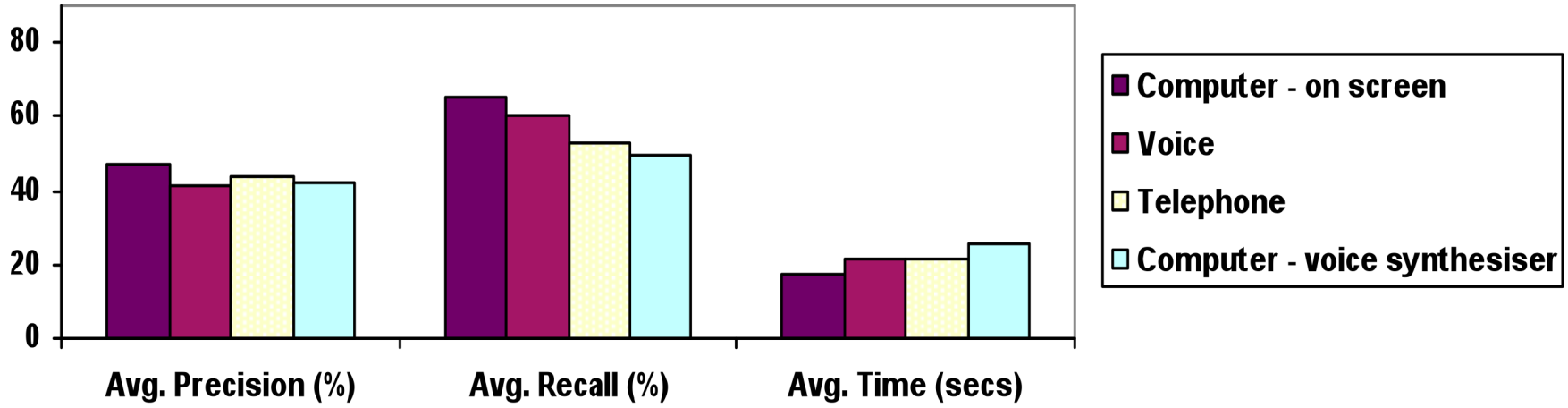
- subject submits a query
- a list of relevant document surrogates is produced
- document surrogates are presented in different modalities (on screen, human voice, human voice on telephone, synthesised voice on telephone)

Take note of effectiveness, speed and ask to fill in a questionnaire at the end

- The real values are not important, but the differences

Experimental Results

Average Precision (P), Recall (R), and Time/ Speed (S) for different modalities of document delivery



Analysis of Results

User perception of relevance is highly influenced by the modality of results presentation

- P varies (but not much) across modalities
- R and S decreases with increasing modality complexity (R and S are related)
- Large across-subjects variations in P, R, S
- Small across-topics variations in P, R, S
- Most participants not happy with the quality of the synthesized voice
- Most participants did not like spoken results presentation: too tiring (fatigue effect on data), difficult memorization of results

Conclusions from exp. results

Presentation of retrieval results using speech is not as effective as results presentation on screen (but not a lot less)

Too much contextual information could be a distraction

New ways of presenting retrieval results *in context* using speech are needed:

- “Highlighting” search terms
- Better speech synthesis systems
- How do we aid results memorization?
- What do we do with multimedia documents?

Conclusions of the talk

IR systems (and Web search engines) are reaching a limit of what can be done with indexing and retrieval models alone

A lot of work on context is currently being carried out, both at academic and commercial level

- Definition and capturing of context
- Click-through data, links analysis, personalisation, …
- Modelling search context, work context, …

“IR in context” is going to bring about a new generation of IR systems … but a lot of research is still needed

Questions