


PART 3 Tutorial © ECIR
Toulouse
6th Apr. 2009

Thomas Mandl
Information Science
University of Hildesheim
mandl@uni-hildesheim.de

Topic Difficulty and Topic Specific Treatment

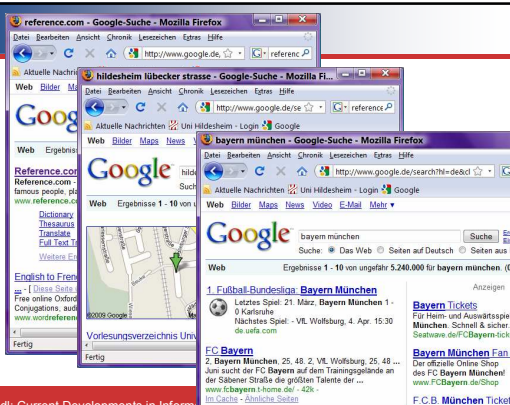


1

Query Classification

- To adapt the user interface (result display)
 - User behavior for navigational and informational queries differs
 - eye tracking study
 - URL or Snippet fixed for a longer time
(Cutrell & Guan 2007)
 - Geographical queries often require a map (Google)

2

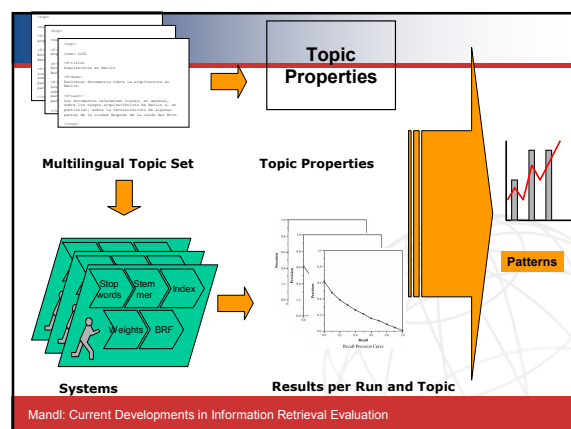
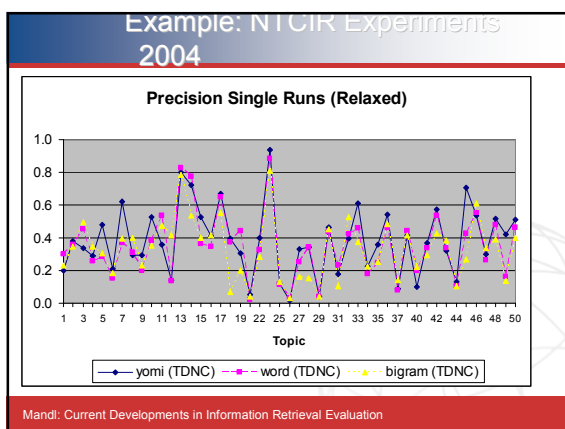


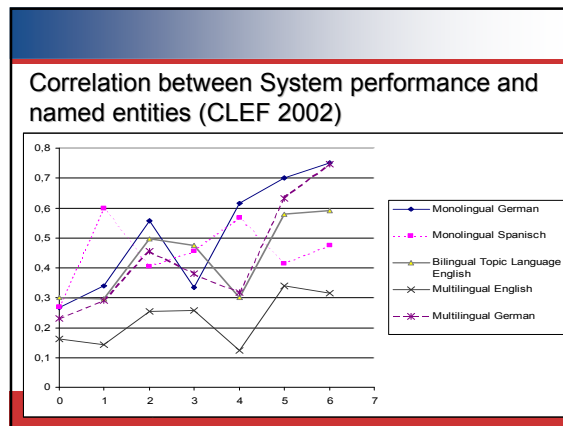
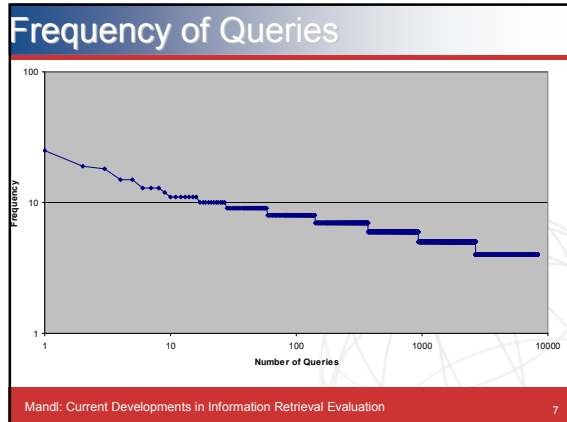
Mandl: Current Developments in Inform...

Query Classification

- Retrieval performance
 - „More work needs to be done on customizing methods for each topic“ *(Harman 2005)*
 - Geographic Queries may require geographic reasoning (GeoCLEF)
 - Difficult queries are solved well by several systems
 - Analysis of errors *(RIA Workshop, Savoy 2007)*

Mandl: Current Developments in Information Retrieval Evaluation





NEs

- Named Entities „simplify“ the Retrieval for the systems
 - Hitherto language-independent phenomenon
 - Statistically significant for several tasks, but not for all

Mandl: Current Developments in Information Retrieval Evaluation

Correlation

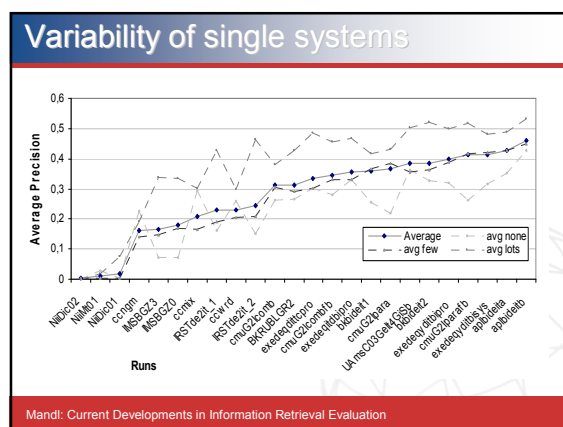
- Results are confirmed via correlation analysis
 - correlation between number of proper names per topic and the *average precision* of all systems for this topic

Mandl: Current Developments in Information Retrieval Evaluation

Correlation between NEs in topic and Retrieval Quality

CLEF Jahr	Task Typ	Topic Sprache	Anzahl runs	Korrelation
2000	Multilingual	English	21	0.26
2001	Bilingual	German	9	0.44
2001	Multilingual	German	5	0.19
2001	Bilingual	English	3	0.20
2001	Multilingual	English	17	-0.34
2002	Bilingual	German	4	0.33
2002	Multilingual	German	4	0.43
2002	Bilingual	English	51	0.40
2002	Monolingual	German	21	0.45
2002	Monolingual	Spanish	28	0.21
2003	Monolingual	German	30	0.37
2003	Monolingual	Spanish	38	0.39
2003	Monolingual	English	11	0.16
2002	Multilingual	English	32	0.29
2003	Bilingual	German	24	0.21
2003	Bilingual	English	8	0.41
2003	Multilingual	English	74	0.31

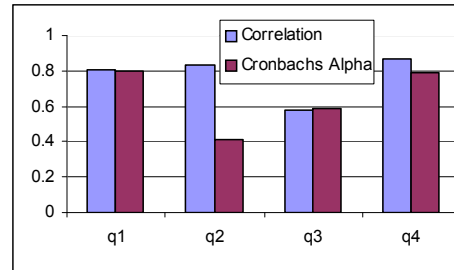
Mandl: Current Developments in Information Retrieval Evaluation



Sub Set Analysis

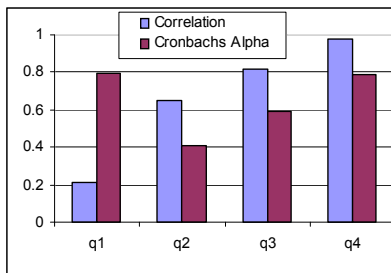
- Sort all 100 topics according to their difficulty (defined as average performance of systems for this topic)
- Split the set into four smaller subsets
- For each quarter
 - Determine a system ranking based on the subset of topics
 - Calculate the correlation to the original ranking (based on the complete set)
- Cronbachs Alpha: Measure for Internal Test reliability
 - Based on
 - Applied to IR (Bodoff & Li @ SIGIR 2007)

Mandl: Current Developments in Information Retrieval Evaluation



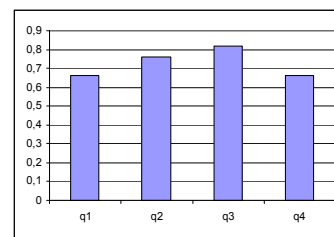
Correlation and Cronbach's Alpha for the Monolingual English sub-task

Mandl: Current Developments in Information Retrieval Evaluation



Correlation and Cronbach's Alpha for the Multilingual sub-task

Mandl: Current Developments in Information Retrieval Evaluation



Multilingual Task: Correlation of the GMAP measure between small sets and full set

Mandl: Current Developments in Information Retrieval Evaluation

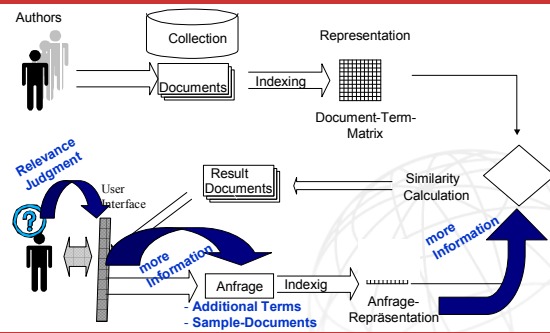
Performance-dependent BRF

- BRF as Standard-technique for optimisation
- Improves normally the MAP, but how does BRF effect the different queries?
- Using terms from the top documents as for query expansion
- Integrating more knowledge into the query

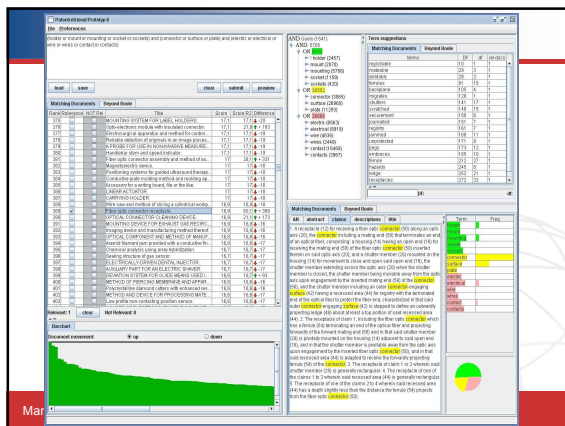
Mandl: Current Developments in Information Retrieval Evaluation

17

IR-Process



Mandl: Current Developments in Information Retrieval Evaluation



Performance-dependent BRF

- Improves normally the MAP, but how does BRF effect the different queries?
 - Negative for the best queries
 - Negative for the worst queries
 - Positive in middle range
- But: you don't know previously, in what of the three categories the query belongs
 - > an estimation of difficulty would be very helpful

Kwok 2005

- Expansion terms gained with BRF are typically not independent
 - Violation of the theoretical assumption on term independence
 - Clustering of terms
 - Expansion with each cluster

(Zaragoza 2009)

Ambiguity

- Clarity respectively ambiguity
 - Measurement of ambiguity of the query by analyzing single terms (Agirre & Mandl)
 - Measurement of coherence in the answer set
 - Distance between the top documents
 - Coherence of top documents (Cronen-Townsend & Croft 2002, He et al. 2008 @ LWA)
 - Distribution of the query terms in the top documents (Amati et al. 2004)
- Query type
 - Ranking proceeding selection according to Query type (Amitay et al. 2003)

Activity 3

- Data Mining on evaluation results
 - explore real results from CLEF 2002
 - Identify difficult and interesting topics!

If you find something really interesting ...