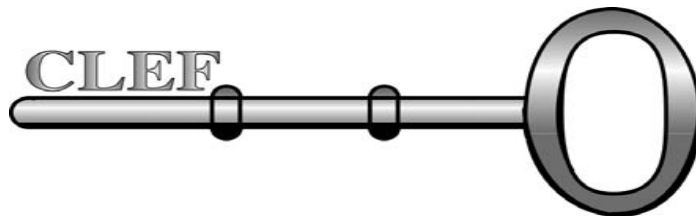# The GIRT task of CLEF 2004

Michael Kluck (IZ)

GIRT Testdatabase

- **German Indexing and Retrieval Testdatabase** containing German social science data

- **1997**   GIRT1   13.000 documents   internal tests
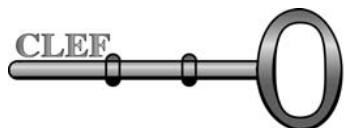
- **1998/99** GIRT2   38.000 documents   TREC

- **2000/02** GIRT3   76.000 documents   CLEF

- **2003/04** GIRT4 151.319 documents **DE**   CLEF

                             151.319 documents **EN**   CLEF

                             >paper on GIRT history in 2003 proceedings

GIRT4 German-English Parallel Corpus

- **151,319 single documents in 2 languages:**
  - Original documents in German
  - Translations into English

- **equal distribution** German/English:
  - **Title**: 1 per document
  - **controlled indexing terms**: ca. 10 per document
  - **Classification text**: ca. 2 per document

- **different distribution** German/English:
  - **Abstract (short text, summary)**
    - DE nearly each document
    - EN 15% of the documents
  - ⇨ reduced textual extent of the English Part
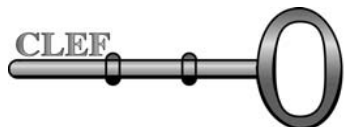
GIRT Tasks in CLEF
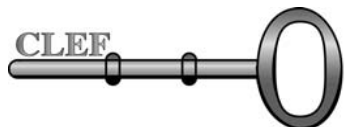
Monolingual Task

- DE Topics -> DE Data
- EN Topics -> EN Data

Bilingual Task

- EN or RU Topics -> DE Data
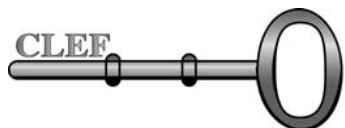- DE or RU Topics -> EN Data

additional instruments

- German-English thesaurus
- German-Russian term list
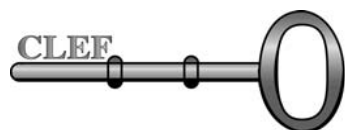
Participation in 2004

4 groups

- **Berkeley, University of California (USA)**
- **University of Hagen (Germany)**
- **IRIT, University of Toulouse (France)**
- **Ricoh (Japan)**

GIRT4 runs by 4 participants

| *Data*	*Topic language*	*Judged runs*	
		2004 2003	
GIRT4 DE	DE	8 (13)	Mono-lingual
GIRT4 EN	EN	7 (4)	15 (17)
GIRT4 DE	EN	6 (1)	Bi-lingual
GIRT4 DE	RU	0 (2)	16 (5)
GIRT4 EN	DE	10 (1)	
GIRT4 EN	RU	0 (1)	
sum		31 (22)	

CLEF

Main Approaches

- **Retrieval: NLP, neural network, logistic regression, probabilistic**

- **Use of IZ thesaurus (once): thesaurus matching, entry vocabulary module**

- **Use of WordNet (twice)**

- **Other: stemming, decompounding, stop words, semantic networks**
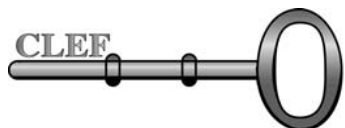
Assessment of GIRT4

- **18,292 (17,031) docs assessed**

  increase of 7 % (65 %) compared to 2003 (2002)

- **Separately German and English part**

  by two assessors (each per language)

- **18,292 hits for 25 topics ➤ 732 hits per topic**

- **16 % judged as actually relevant (2003: 20 %)**

 min. 0.6 % (n= 3), max. 57 % (n= 232) per topic

CLEF
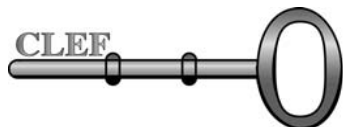
Assessment in detail

- **The search results in the different language parts have not been congruent**

- **Only for 2 topics an identical number of hits**

  German / English

- **10% of hits show pairs of corresponding relevant documents**

  German / English

- **16 % of hits show pairs of non-relevant documents German / English**

- **Overlap German / English results is 26 %**
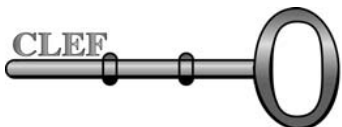
Comparison of judgments

- **in 1.86% cases (n= 340) divergent judgments for corresponding DE/EN**

➤ **re-assessment of these documents**

  ➤ **35% real mistakes (n=120)**

  ➤ **44% new interpretation of relevance (n=149)**

  ➤ **21% too short text for equal judgment (n=71)**

➤ **Re-assessed docs included in official figures**

➤ **further analysis in proceedings 2004**

CLEF
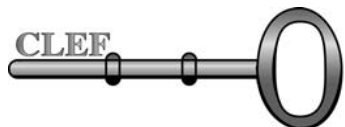
Analysis of individual results

- **Concordance of document numbers of identical documents DE/EN available**

- **For specific topics: participants and IZ work together in analysing**

  - **failure of delivering the corresponding document pairs DE/EN for bi-lingual runs**

  - **failure of delivering known relevant documents for mono-lingual runs**

CLEF