

The GIRT Task at CLEF 2002

A short introduction

Michael Kluck

IZ Sozialwissenschaften Bonn / Humboldt-University Berlin

kluck@bonn.iz-soz.de

GIRT task

- Perform searches against the domain-specific collection GIRT as
 - Monolingual task: German topic against German GIRT data
 - Bilingual task: English or Russian topics against German GIRT data (other topics language are also possible if independent translation of topics is provided)

Allowed GIRT fields

- TITLE
- TITLE-ENG
- TEXT
- TEXT-ENG
- CONTROLLED TERM
- FREE TERM

Runs might have made use of the controlled terms and the free terms and of the provided German-English Thesaurus or German-Russian wordlist

GIRT assessment

- 25 topics (German, English, Russian)
- 10.323 documents in the pool (413 docs/topic)
- 961 relevant documents in the pool (38 docs/topic)
- One topic without relevant answers
- Between 0 or 0.5% and 32,1% relevant documents per topic (average 9,6%)

GIRT participation in CLEF 2002

5 groups (used formats; used languages)

- Univ. Amsterdam (TD; DE, EN)
- Univ. Berkeley (TD, TDN; DE, EN, RU)
- Univ. Dortmund (TD; DE)
- Univ. Hildesheim (TD; DE)
- Xerox Xrce (T; EN)

Specific challenges

- Social science terminology
- More text fields available
 - TITLE; TEXT; CONTROLLED TERM (= indexing terms from the social science thesaurus); FREE TERM (= additional intellectually chosen terms)
 - Possibility to use a part of the collection as parallel English collection by using the English text fields and titles (although they are “only” translations from German, and not for all documents available)
 - Make use of indexing terms of the documents
- German-English Thesaurus, German-Russian wordlist

Outlook 2003

- Idea: Combination with the social science part of the FRANCIS database of INIST -> two corpora
- French, German, English, Russian topics
- Reformatting GIRT by making a clear parallel “English” version of the GIRT Data available (translated title and text fields)
- Enlarging the GIRT data