

Exeter at CLEF 2003: Cross-Language Spoken Document Retrieval Experiments

Adenike M. Lam-Adesina

Department of Computer Science, University of Exeter, U.K.

Gareth J. F. Jones

School of Computing, Dublin City University, Ireland

CLIR Strategy

- Retrieval using the City University research distribution of the Okapi system.
 - Around 260 stop words removed from the texts, Porter stemming applied, small set of standard synonyms used.
- Okapi augmented with summary-based pseudo relevance feedback (Lam-Adesina & Jones SIGIR 2001).
- Pilot searching incorporating 20,000 parallel text documents from TDT-2 collection.

Monolingual German Runs

	English	French	German	Italian	Spanish
Baseline	0.311	0.227	0.203	0.231	0.250
TC $ow(i)$, TC $cfw(i)$	0.382	0.281	0.270	0.279	0.292
% chg.	+22.8%	+23.8%	+33.0%	+20.7%	+16.8%
Rel. Ret.	1795	1558	1498	1638	1641
CC $ow(i)$, TC $cfw(i)$	0.364	0.283	0.274	0.299	0.304
% chg.	+17.0%	+24.7%	+34.9%	+29.4%	+21.6%
Rel. Ret.	1824	1618	1541	1684	1720

Systran Topic translation.

Further Work

- Complete experiments using pilot collections for topic expansion and term weighting.
 - TDT-2 text documents.
 - TREC text documents.
 - Other data?