Interactive Translingual Searching Using Document Expansion

Bonnie Dorr, Gina-Anne Levow, Douglas Oard, Philip Resnik, Clara Cabezas, Okan Kolak and Jianqiang Wang
University of Maryland
http://www.umiacs.umd.edu/labs/CLIP/

Web browser forms interface

English Query Interface

Comparable English Documents

English Translations

n-best Gloss Translation

Chinese Documents

Ordered by Brown Corpus frequency
1-best or 2-best for retrieval
3-best for examination

Rank list with translated titles

Selection Interface

Document Expansion

Select

Translate Examination Interface

1-best Machine Translation

System Professional 2.0

Interactive

Differential emphasis

Scalable

Detection Error Tradeoff (lower is better)

- Expansion is done at indexing time
- No significant impact on search time
- Comparable documents often exist
- A natural part of many multilingual applications
- Builds on existing retrieval systems
- The most complex components are reusable

- 3-best Glossed English
- Random Choice

Classification Smooth (Error Rate)

Noise Bayes Classifier

TDT-3 Chinese newsreader test

No expansion

16-best document translation

20-best document translation