

ImageCLEFphoto

OPENOFFICE BASED THESAURUS

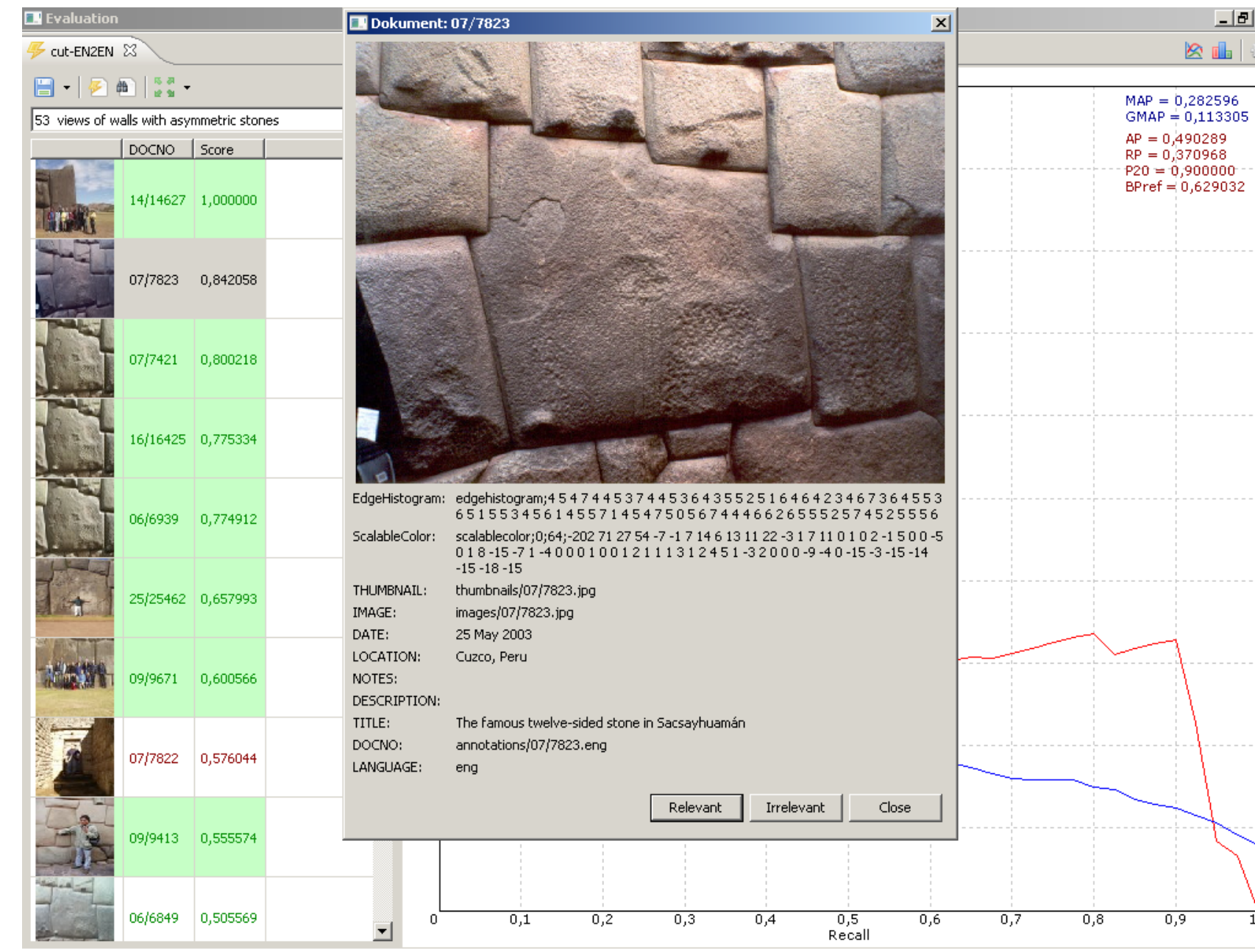
- Characteristics:**
- based on the free available thesauri of the OpenOffice.org project
 - highly customisable by use of different analysers

- Procedure:**
- > create Lucene index with an appropriate analyser:
 - the original word tokenized, indexed, but not stored
 - all assigned words only stored within the same Lucene document
 - > use the whole topic (title) as query to this index
 - > inject results into the original query with a custom weight (i.e. 0.5)

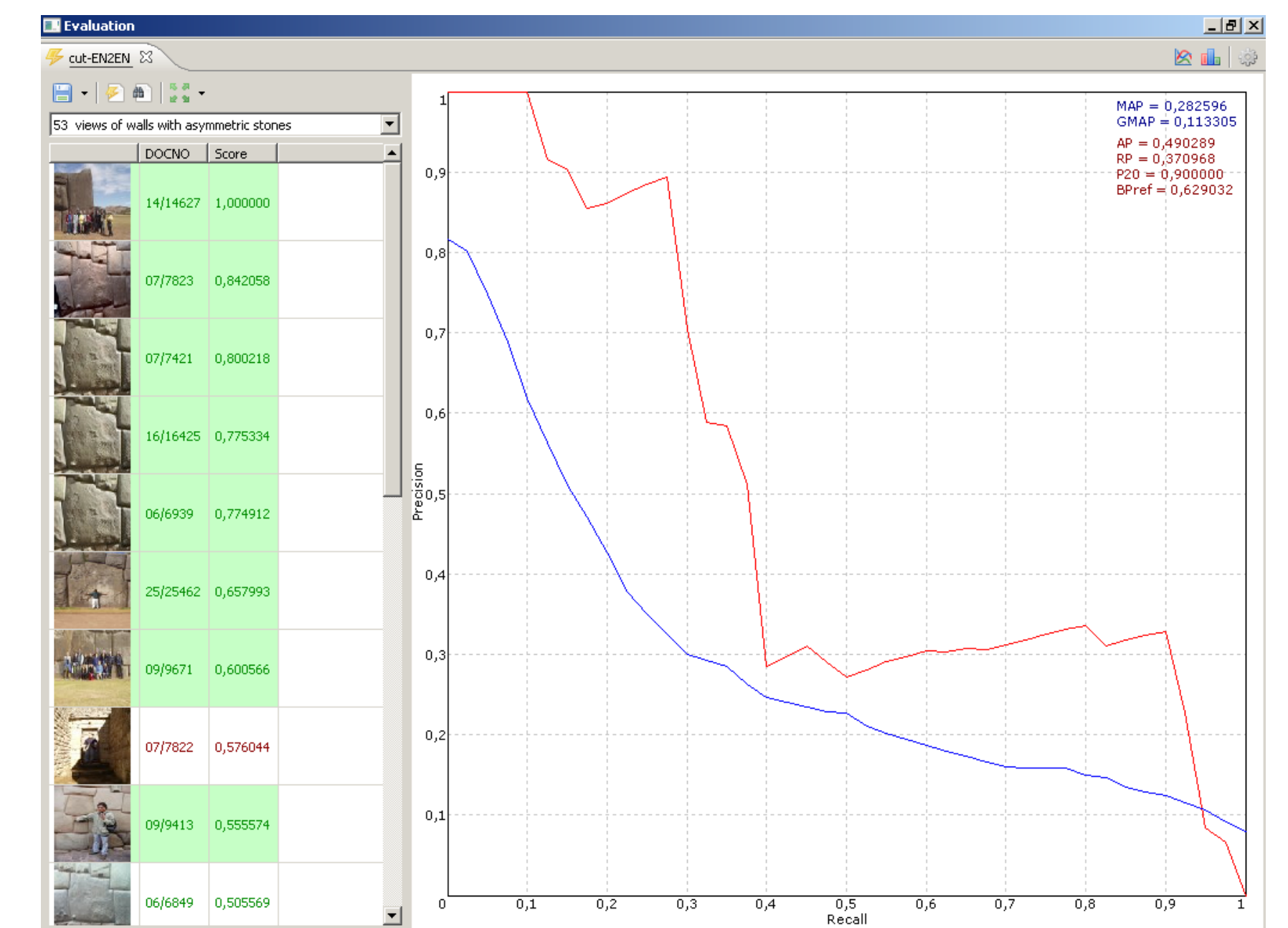
QUERY EXPANSION

- Characteristics:**
- tightly bound to the Lucene index architecture
 - independent of index structure

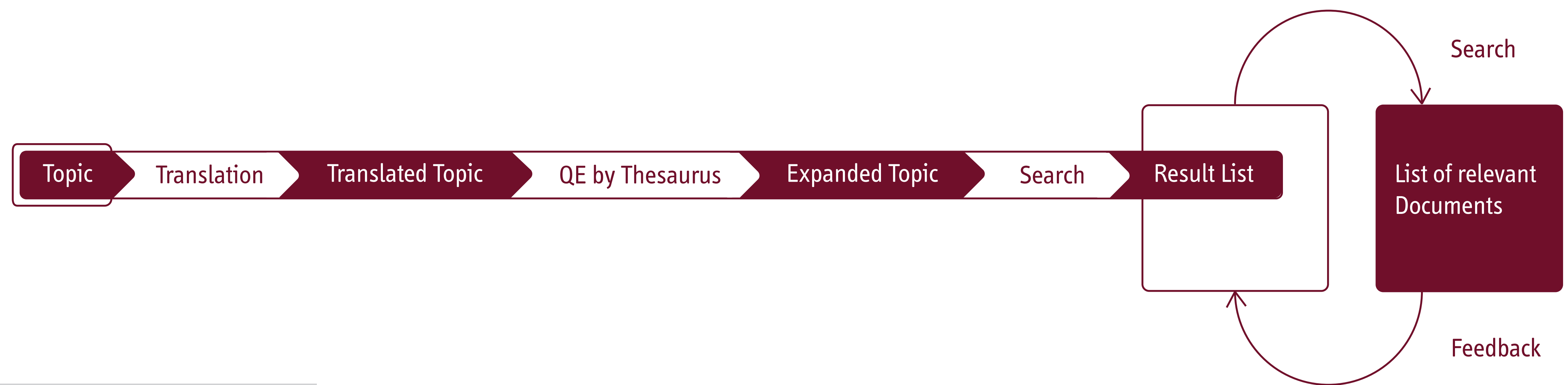
- Procedure:**
- > fetch and count all terms (in every field) from relevant documents
 - > apply a custom threshold to extract the important terms
 - > inject these terms into the Lucene query
 - > repeat the search



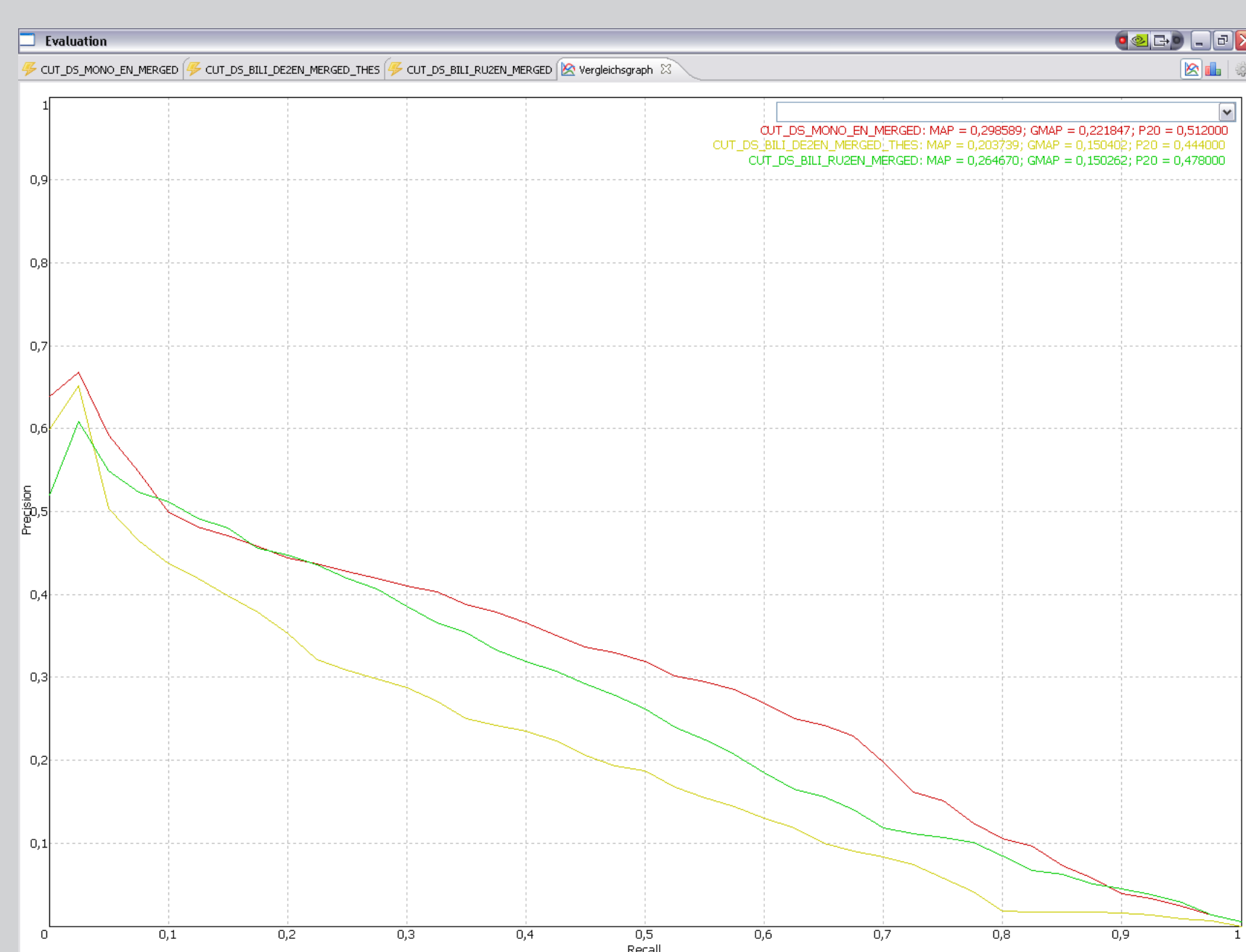
Detail view of a single document from the result list with annotations also visible



Evaluation window with resultlist and recall precisiongraph for the topic and the whole run

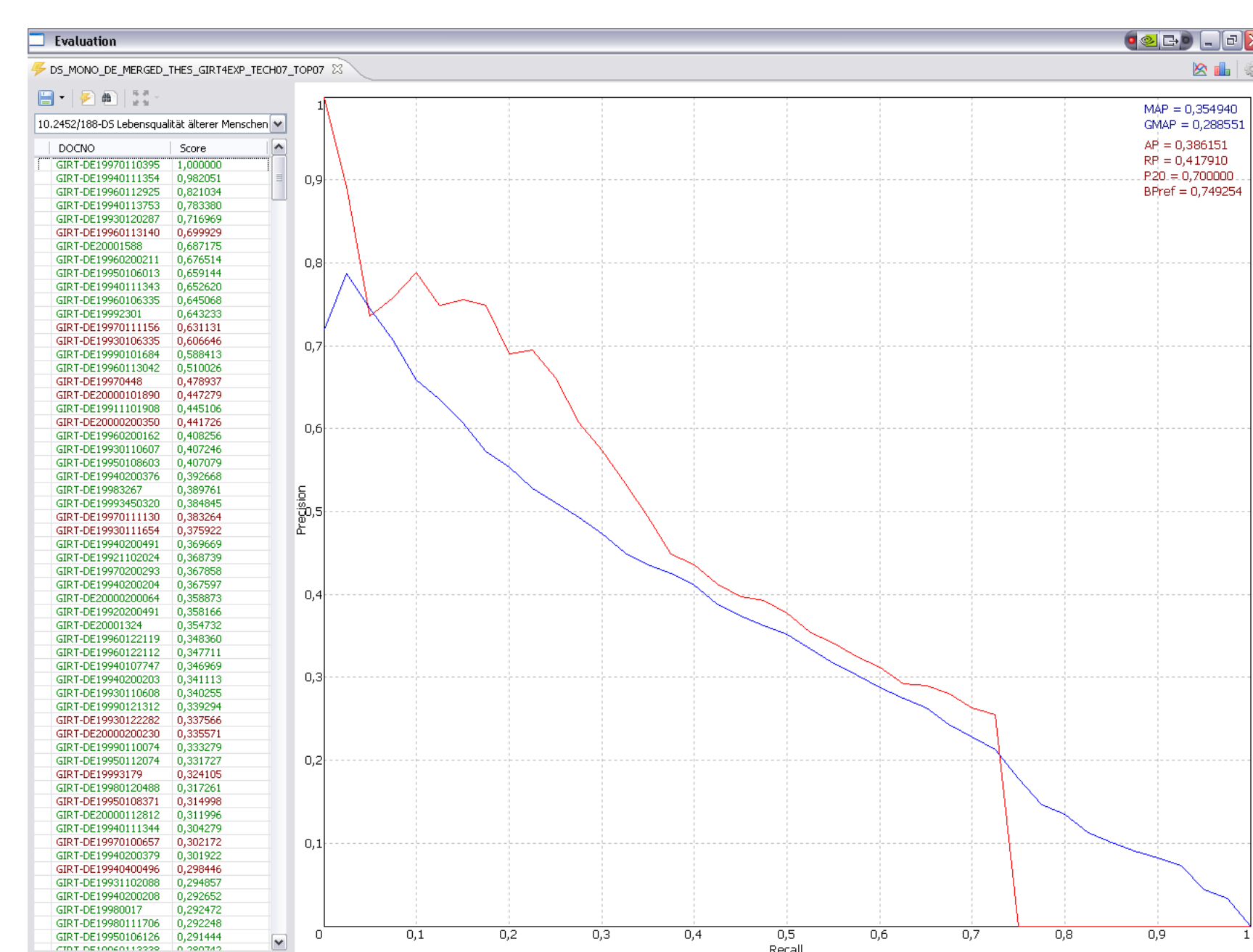


Domain-Specific Cross Language Retrieval



CROSS-LINGUAL ENGLISH

- translation: bilingual thesaurus + Google translate BETA
 index: structured and plain
 data fusion: z-score operator
 results: close to monolingual results
 Russian to English performs best

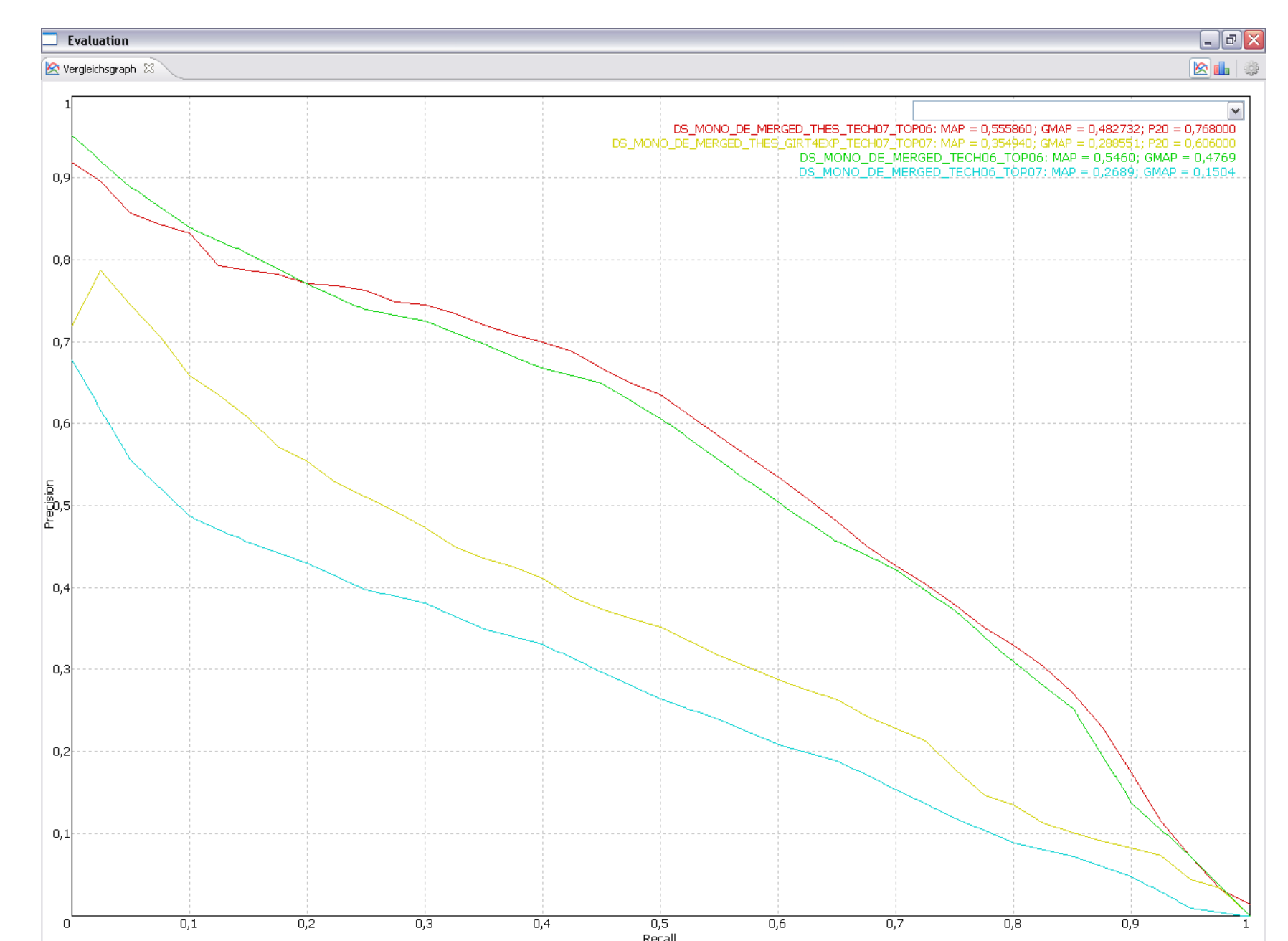


EVALUATION

- graphical comparison: average precision vs. mean average precision
 traditional values: (G)MAP, BPref, R-Precision, P@20
 access to document contents of retrieved result list

DISTINCTIONS IN THE TOPICS

- 2006:** mainly general topics
 > i.e. „Poverty“, „Street urchins“, „Risk behavior“
 > many documents satisfy the information need > task more or less easy
- 2007:** more restrict nature of topics
 > i.e. „Lifestyle and environmental behavior“, „Economic elites in Eastern Europe and Russia“
 > fewer documents satisfy the information need > more difficult task



DISTINCTIONS IN THE SYSTEM

- 2006:** Text Analyzer: Decomposing + Suffix Removal
 Index: Data Collection Structure
 Topic Preprocessing: Term Frequency Filter
 Query Expansion: Local Document Clustering + Top-10 Pseudo Relevance Feedback
- 2007:** Text Analyzer: Suffix Removal
 Index: Data Collection Structure + Plain
 Topic Preprocessing: Term Frequency Filter + Thesaurus/GIRT4 Keywords
 Query Expansion: Top-10 Pseudo Relevance Feedback