

## Ad-Hoc TEL Task Retrieval on Multilingual Data Collections

### APPROACH AND RESOURCES

#### COMPONENTS

- Xtrieval Framework
- IR Models: BB2  
TF.IDF  
BB2 + TF.IDF
- Stemming: EN (Snowball + Krovetz)  
DE (Snowball)  
FR (Snowball)
- Standard Stoplists for EN, DE and FR
- Translation: Google AJAX Language API

#### EXPERIMENTAL SETUP

- Merging for EN and DE Runs
- Pseudo-Relevance Feedback

#### TACKLING THE MULTILINGUALITY

- Generating Multiple Indexes
- Using Top-4 Occuring Languages
- Language Distribution as Weights

### EXPERIMENTAL RESULTS

#### MONOLINGUAL

experiment ID	IR Core / Model	lang	MAP
cut1	Terrier (BB2)	DE	0.2602
cut2	Lucene (TF.IDF)	DE	0.2641
cut3	Merged	DE	0.2789
cut4++	Terrier (BB2)	DE	0.2713
cut9	Terrier (BB2)	EN	0.3864
cut10	Lucene (TF.IDF)	EN	0.3672
cut11	Merged	EN	0.4071
cut12++	Terrier (BB2)	EN	0.3914
cut17	Terrier (BB2)	FR	0.2470
cut18	Lucene (TF.IDF)	FR	0.2399
cut19	Merged	FR	0.2583
cut20++	Terrier (BB2)	FR	0.2540

#### BILINGUAL

experiment ID	IR Core / Model	lang	MAP
cut5	Merged	EN->DE	0.2583
cut7++	Terrier (BB2)	EN->DE	0.2580
cut13	Merged	DE->EN	0.4046
cut15++	Terrier (BB2)	DE->EN	0.3427
cut21	Merged	EN->FR	0.2320
cut24++	Terrier (BB2)	DE->FR	0.2557

## Grid@CLEF

## Combination of Different IR Models

### APPROACH AND RESOURCES

#### COMPONENTS

- Xtrieval Framework
- CIRCO Framework
- IR Models: BM25  
TF.IDF  
TF.IDF + BM25
- Stemming: EN (Snowball + Krovetz)  
DE (Snowball + N-Gram)  
FR (Snowball + Savoy)
- Standard Stoplists for EN, DE and FR

#### EXPERIMENTAL SETUP

- Comparing + Merging IR Models
- Pseudo-Relevance Feedback

#### LESSONS LEARNED

- Very Large XML-Documents!
- Only Chains of Processing Possible!
- Exchange of Data vs. Processing Code?

### EXPERIMENTAL RESULTS

#### MONOLINGUAL-DE

experiment ID	IR Model / Core	Stemmer	MAP
cut_de_1	Lucene (TF.IDF)	Snowball	0.4196
cut_de_2	Terrier (BM25)	Snowball	0.4355
cut_de_3	Lucene (TF.IDF)	N-Gram	0.4267
cut_de_4	Terrier (BM25)	N-Gram	0.4678
cut_de_5	Merged	both	0.4864

#### MONOLINGUAL-EN

experiment ID	IR Model / Core	Stemmer	MAP
cut_en_1	Lucene (TF.IDF)	Snowball	0.5067
cut_en_2	Terrier (BM25)	Snowball	0.4926
cut_en_3	Lucene (TF.IDF)	Krovetz	0.4937
cut_en_4	Terrier (BM25)	Krovetz	0.4859
cut_en_5	Merged	both	0.5446

#### MONOLINGUAL-FR

experiment ID	IR Model / Core	Stemmer	MAP
cut_fr_3*	Lucene (TF.IDF)	Snowball	0.4483
cut_fr_1	Terrier (BM25)	Snowball	0.4538
cut_fr_5	Lucene (TF.IDF)	Savoy	0.4434
cut_fr_2	Terrier (BM25)	Savoy	0.4795
cut_fr_4	Merged	both	0.4942

## ImageCLEFPhoto Task Baseline Text Retrieval Experiments

### USED RESOURCES

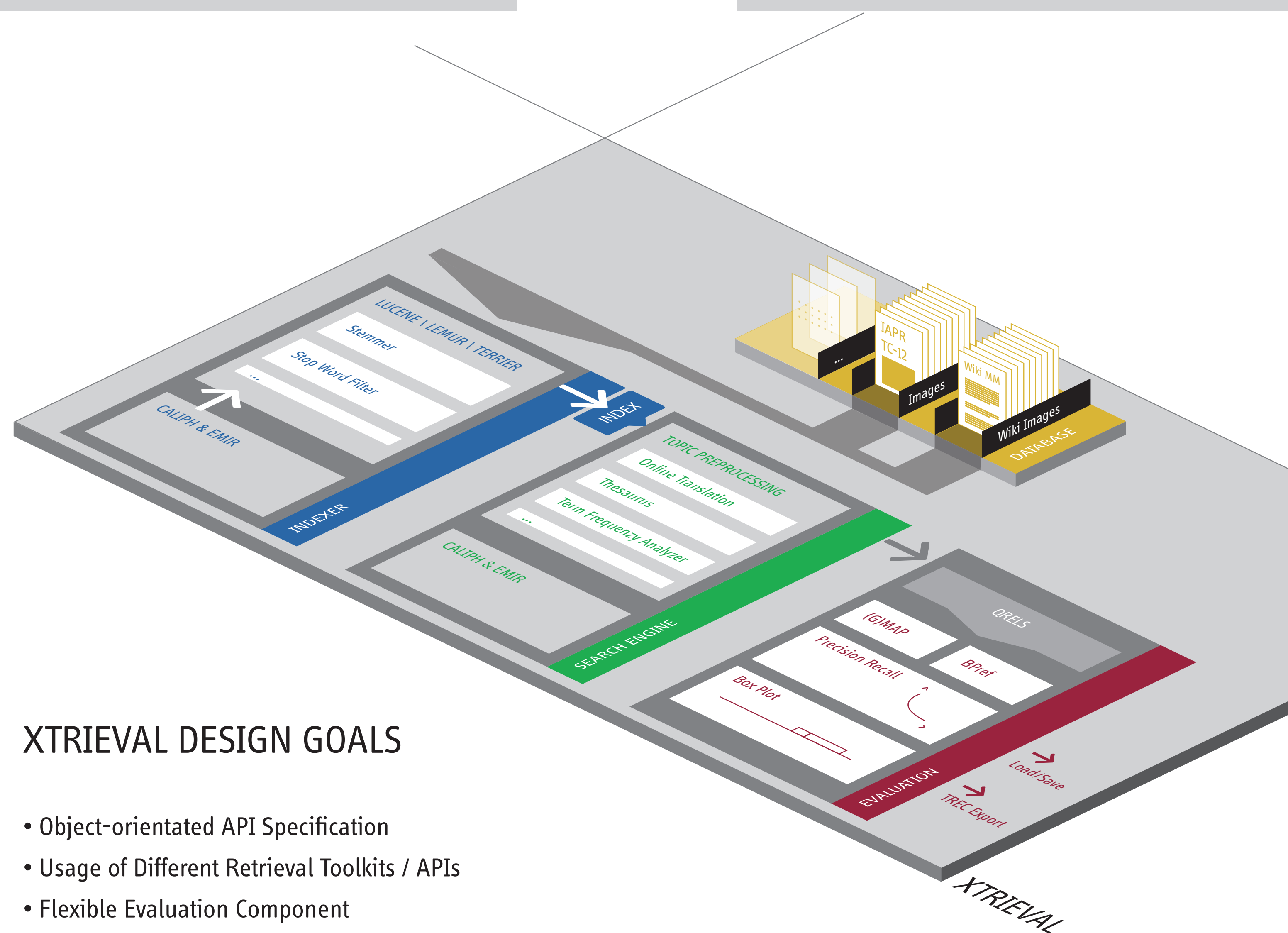
- Xtrieval Framework for Text Retrieval
- Lucene (TF.IDF) and Terrier (Okapi BM25) Models
- Query Tags: Titles + Cluster Titles

### EXPERIMENTAL RESULTS

Experiment ID	Model	MAP
CUT2_T_TXT	Terrier (BM25)	0.5041
CUT2_TCT_TXT	Terrier (BM25)	0.4821
CUT1_T_TXT	Lucene (TF.IDF)	0.4468
CUT1_TCT_TXT	Lucene (TF.IDF)	0.4386

### CONCLUSIONS

- Average Performance in Terms of F-Measure
- Strong Results in Terms of MAP
- BM25 >> TF.IDF
- Titles Only > Titles + Cluster Titles



### XTRIEVAL DESIGN GOALS

- Object-orientated API Specification
- Usage of Different Retrieval Toolkits / APIs
- Flexible Evaluation Component
- Simple but Flexible Configuration

### XTRIEVAL FRAMEWORK DESCRIPTION

The basic idea of the Xtrieval Framework is to provide interfaces to combine different state-of-the-art text retrieval techniques on the one hand and to integrate and evaluate new methods for multimedia retrieval on the other hand.

## VideoCLEF Classification Task ASR-based Video Classification as IR Task

### USED RESOURCES

- Xtrieval Framework with Lucene (TF.IDF)
  - OpenOffice English Thesaurus for QE
  - Google's language API for Translation

### EXPERIMENTAL RESULTS

experiment ID	Avg. Recall	MAP
cut1_l1_base_asr	0.0101	0.0067
cut2_l1_qe_asr	0.0958	0.1010
cut3_l1_base_meta	0.2814	0.2586
cut_4_l1_qe_meta	0.3568	0.2531
cut_5_l3_qe_meta	0.5379	0.3813

### CONCLUSIONS

- ASR + Metadata Improves Performance
- Best MAP with unlimited Docs / Label
  - Highest Correct Classification Rate with 1 Doc / Label