

## Ad-Hoc Track On-Line Processing Experiments

### APPROACH AND RESOURCES

#### COMPONENTS

- Xtrieval Framework with Lucene core
- Stemming: EN (Snowball + Krovetz)  
DE (Snowball + n-Gram)  
FR (Snowball)
- Standard stopwords lists for EN, DE and FR
- Translation: Google AJAX language API

#### EXPERIMENTAL SETUP

- Live experiments because of wrong DOCID
- Data fusion for EN and DE runs
- Top-10 pseudo-relevance feedback

#### TACKLING THE MULTILINGUALITY

- Generating multilingual queries
- Top-10 occurring languages
- Using language distributions as weights

### EXPERIMENTAL RESULTS

#### MONOLINGUAL

experiment ID	lang	MAP
cut_merged_simple	DE	0.2109
cut_multi0_wx++	DE	0.1795
cut_multi0_wl++	DE	0.1113
cut_multi_wl-x++	DE	0.1073
cut_merged_simple	EN	0.3562
cut_multi0_wx++	EN	0.2484
cut_multi0_wl-x++	EN	0.1620
cut_multi0_wl++	EN	0.1453
cut_merged_simple	FR	0.1981
cut_multi0_wx++	FR	0.1629
cut_multi0_wl-x++	FR	0.0929
cut_multi0_wl++	FR	0.0915

#### BILINGUAL

experiment ID	lang	MAP
cut_merged_simple_en2de	EN->DE	0.1852
cut_merged_simple_multi0_wl	EN->DE	0.1126
cut_merged_simple_de2en	DE->EN	0.3416
cut_merged_simple_multi0_wl	DE->EN	0.1475
cut_merged_simple	EN->FR	0.1981
cut_merged_simple_multi0_wl	EN->FR	0.1270

## Domain-Specific Track Combining Different Stemmers for CLIR

### APPROACH AND RESOURCES

#### COMPONENTS

- Xtrieval Framework with Lucene core
- Stemming: EN (Snowball + Krovetz)  
DE (Snowball + n-Gram)  
RU (Suffix Removal)
- Standard stopwords lists for EN, DE and RU
- Provided thesauri for QE
- Translation: Google AJAX language API

#### EXPERIMENTAL SETUP

- Data fusion for EN, DE and RU runs
- Top-k pseudo-relevance feedback
- Thesaurus-based QE

#### MULTILINGUAL SETUP

- Translation to all target languages
- Data fusion in retrieval stage

### EXPERIMENTAL RESULTS

#### MONOLINGUAL

experiment ID	lang	MAP
cut_merged	DE	0.4367
cut_merged_thes	DE	0.4359
cut_merged	EN	0.3891
cut_merged_thes	EN	0.3869
cut_merged	RU	0.0955

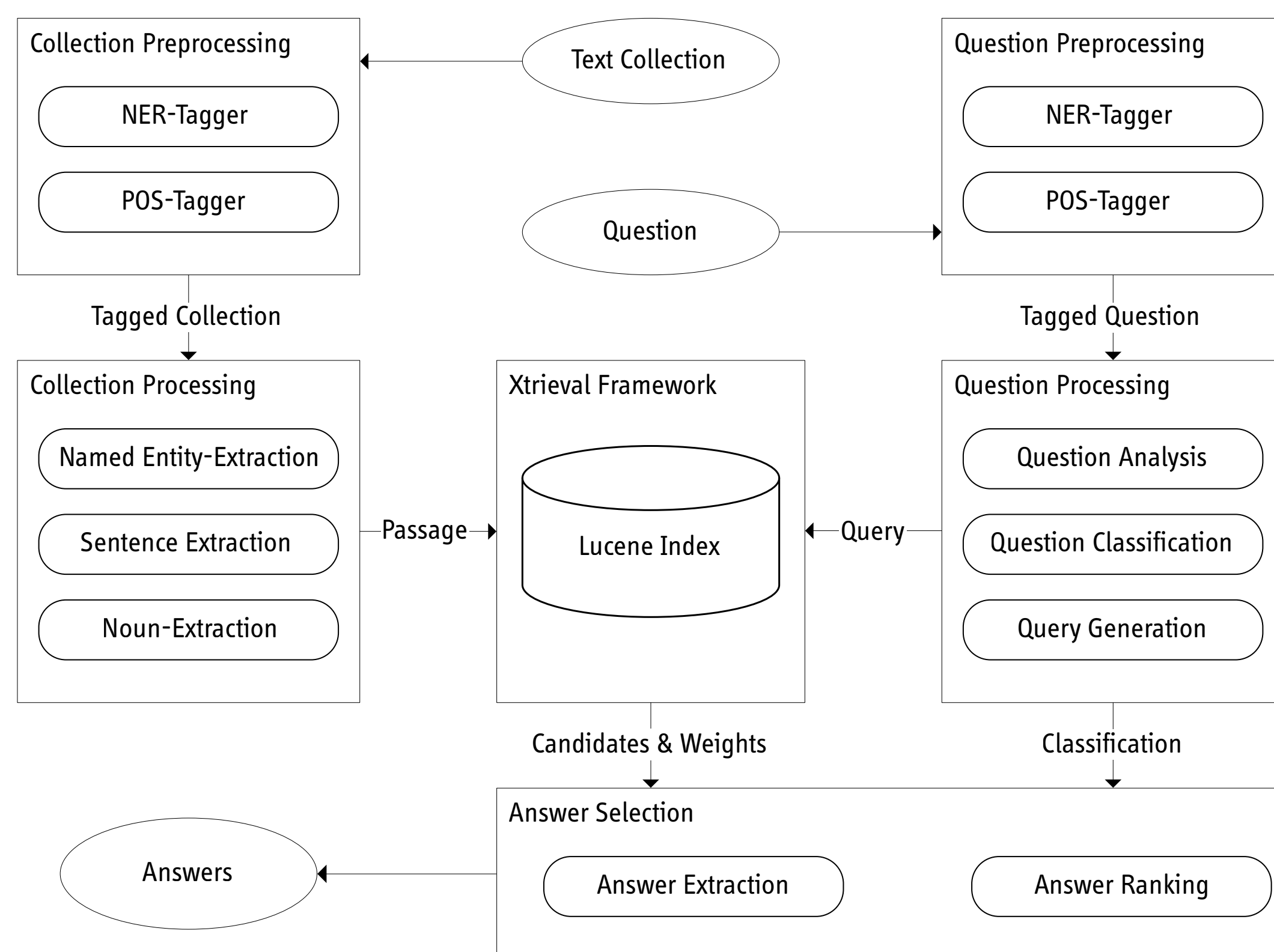
#### BILINGUAL

experiment ID	lang	MAP
cut_merged_en2de	EN->DE	0.3702
cut_merged_en2de_thes	EN->DE	0.3554
cut_merged_ru2en	RU->EN	0.3385
cut_merged_de2en	DE->EN	0.3363
cut_merged_en2ru	EN->RU	0.0822
cut_merged_de2ru	DE->RU	0.0681

#### MULTILINGUAL

experiment ID	lang	MAP
cut_merged_en2x	EN->X	0.2751
cut_merged_ru2x	RU->X	0.2357

## QAsT Task QA on Manual Speech Transcripts



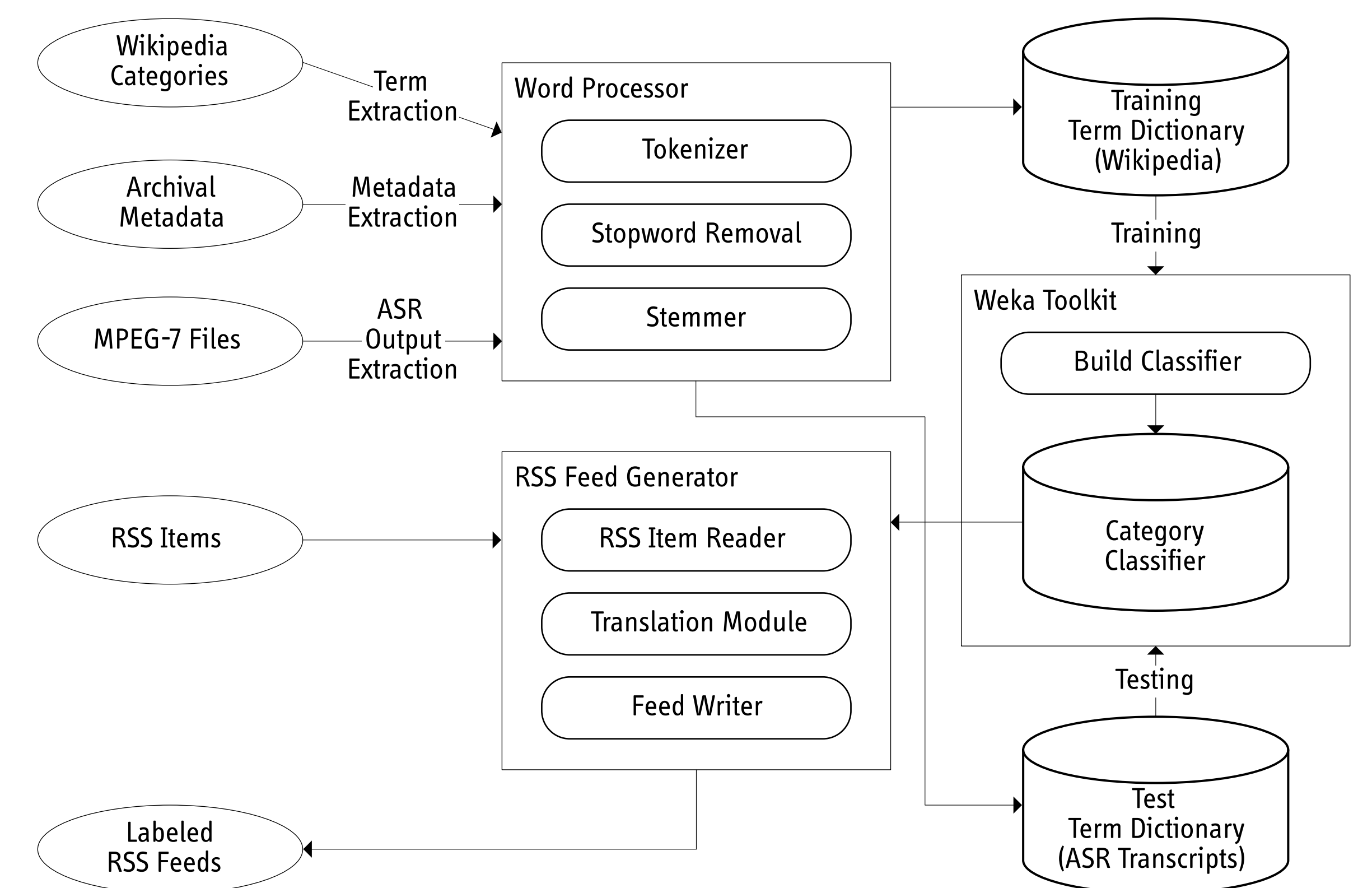
### USED RESOURCES

- Stanford named entity recognizer (SNER)
- Conditional random fields POS tagger
- Hand-crafted patterns for question classification

### EXPERIMENTAL RESULTS

run ID	MRR	ACC
cut1 t1a	0.16	16.0%
cut2 t1a	0.20	17.0%
cut1 t4a	0.21	21.0%
cut2 t4a	0.22	21.0%

## VideoCLEF Track ASR Classification with Wikipedia



### USED RESOURCES

- Dutch / English Wikipedia (IWPL) as training data
- Weka toolkit for classification
- Google's language API for translation

### EXPERIMENTAL RESULTS

run ID	precision	recall
cut_clr1	0.15	0.14
cut_clr2	0.10	0.12
cut_c2r1	0.13	0.12
cut_c2r2	0.12	0.14

#### Authors

Prof. Dr. Maximilian Eibl  
Dipl.-Inf. Jens Kürsten  
Dipl.-Inf. Thomas Wilhelm  
Daniel Richter, Holger Kundisch

#### Affiliation

Chemnitz University of Technology  
Faculty of Computer Science  
Dept. Media and Computer Science  
09107 Chemnitz / Germany

#### Acknowledgments

This work was partially accomplished in conjunction with the project sachsMedia, which is funded by the Entrepreneurial Regions program of the German Federal Ministry of Education and Research from April 2007 to March 2012.