A Modified Information Retrieval Approach to Produce Answer Candidates for Question Answering

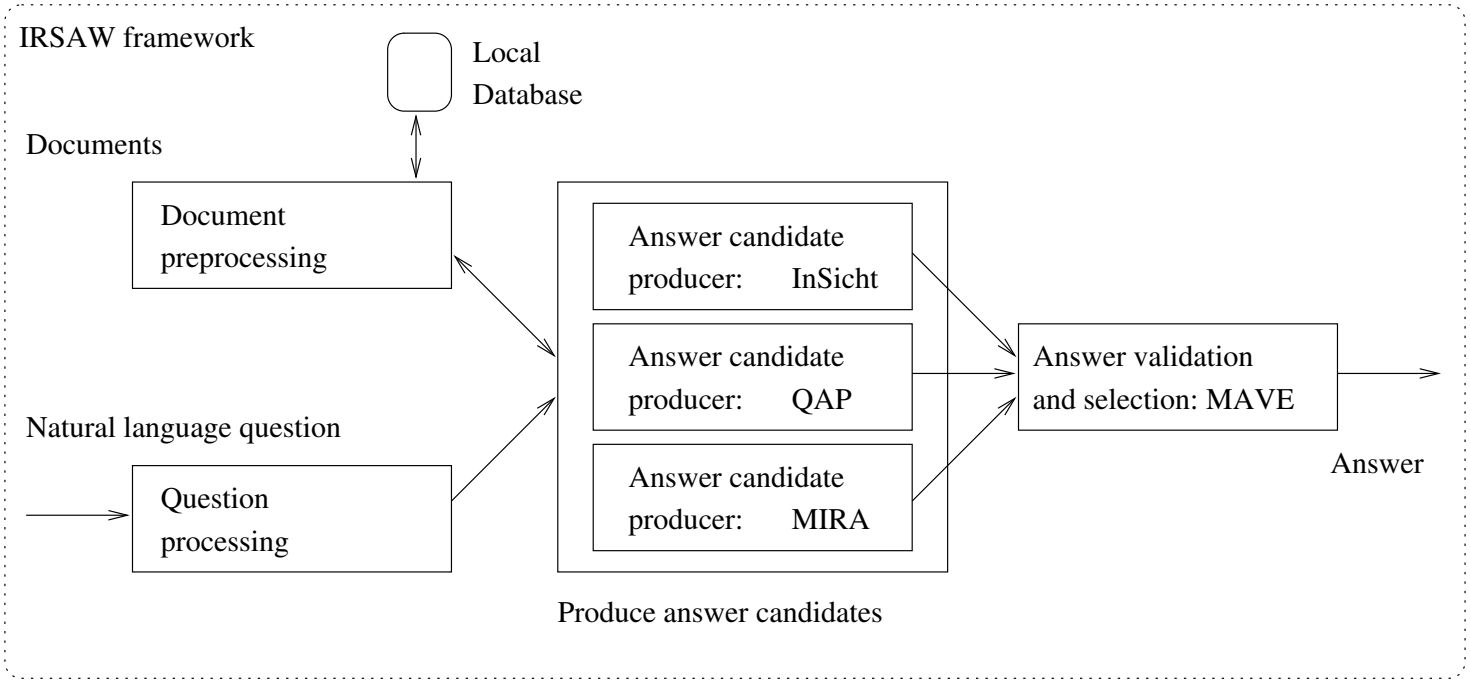
Johannes Leveling

Intelligent Information and Communication Systems (IICS)
University of Hagen (FernUniversität in Hagen)
58084 Hagen, Germany
johannes.leveling@fernuni-hagen.de

LWA 2007 Workshop, Halle (Saale), Germany

Outline

IRSAW question answering framework



IRSAW: ***I**ntelligent Information **R**etrieval on the Basis of a **S**emantically **A**nnotated **W**eb*

funded by the DFG (Deutsche Forschungsgemeinschaft)

Question answering phases

1. Process document collection
2. Preprocess question
   ($\Leftarrow$ Natural language question)
3. Retrieve text segments
4. Match document and question representations
5. Return answer candidates
6. Merge and validate answer candidates
   ($\Rightarrow$ Answer)

Embedding of MIRA in IRSAW

- Employ different modules to produce data streams containing answer candidates:
  - InSicht (Matching semantic network representations, Hartrumpf and Leveling (2007))
  - QAP (Question Answering by Pattern matching, Leveling (2006)), and
  - MIRA (Modified Information Retrieval Approach)
- Use different methods to produce answer streams to increase recall and robustness
- Merge, rank, logically validate answer candidates and select best answer, (MAVE, Glöckner et al. (2007))

MIRA

- Shallow question answering
- Expected answer type (EAT) of question determined by Bayesian classifier: *PERSON*, *SUBSTANCE*, ...
- Manually annotated corpus with EAT tags (e.g. *PERSON*) and subclasses (e.g. person-first person-last)
- TüBa-D/Z newspaper corpus (Tübingen Treebank of Written German; `http://www.sfs.uni-tuebingen.de/en_tuebadz.shtml`), approximately 470,000 words

Expected answer types (1/3)

- Question (German): *Wer wurde 1948 erster Ministerpräsident Israels?*
- Question (English): Who became the first Prime minister of Israel in 1948?
- EAT: *PERSON*
- Answer string:
  David ben Gurion
- Tag sequence:
  person-first person-part person-last

Expected answer types (2/3)

- Question (German): *In welchem Jahr endete offiziell die Besetzung Deutschlands?*
- Question (English): In what year did the occupation of Germany officially end?
- EAT: *TIME*
- Answer string:
  im Jahr 1955
- Tag sequence:
  prep year num-card

Expected answer types (3/3)

- Question (German): *Wie wird der Ebolavirus übertragen?*
- Question (English): How is the Ebola virus transmitted?
- EAT: *OTHER*
- Answer string: (Übertragen werden die Ebolaviren durch direkten Körperkontakt und bei Kontakt mit Körperausscheidungen infizierter Personen per Kontaktinfektion bzw. Schmierinfektion.)
- Tag sequence:
  – (other entity type → answer not found!)

EAT frequency in annotated TüBa-D/Z

Name class	Corpus frequency
LOCATION	8,274
PERSON	14,527
ORGANIZATION	7,148
TIME	14,524
MEASURE	895
SUBSTANCE	293
OTHER	2,987

EAT subclass frequency in annotated TüBa-D/Z

LOCATION	Subclass frequency
city	3,717
country	1,955
region	926
street	613
state	370
other	206
building	195
streetno	124
river	85
island	55
sea	17
mountain	11

Tagging with subclasses

Token	EAT	Subclass
Vor	*TIME*	prep
25	*TIME*	num-card
Jahren	*TIME*	year
betrat	–	
Neil	*PERSON*	person-first
Armstrong	*PERSON*	person-last
als	–	
erster	–	
Mensch	–	
den	–	
Mond	*LOCATION*	other
,	–	
doch	–	
heute	*TIME*	deictic
stagniert	–	
die	–	
bemannte	–	
Raumfahrt	–	
.	–	

MAVE - MultiNet-based Answer Verification

- Validate answer candidates
- Test logical validity of answer candidate by using
  a) inferences, entailments
  b) heuristic quality indicators (fallback strategy)
- Select most trusted answer

Evaluation results (1/3)

Performance results for InSicht, QAP, and MIRA based on questions from QA@CLEF data from 2004 to 2006

System	# Candidates	Coverage	# Correct	Precision
InSicht	1,212	226/600	625	51.6%
QAP	2,562	114/600	1,190	46.6%
MIRA	14,946	520/600	1,738	11.6%

Evaluation results (2/3)

Performance results including answer selection by MAVE based on questions from QA@CLEF data from 2004 to 2006

Run	# Correct	# Inexact	# Wrong
InSicht+Mira+QAP	247.4	15.8	307.8
InSicht+Mira+QAP (opt.)	305.0	17.0	249.0

Evaluation results (3/3)

Results for MIRA answer candidates for QA@CLEF data from 2003 to 2006

	top-*N*			
	N=50	N=30	N=10	N=5
# Correct (2006)	798	615	215	95
# Inexact (2006)	56	53	20	12
# Wrong (2006)	4,436	3,421	1,360	722
# Correct (2003–2006)	1,864	1,503	609	263
# Inexact (2003–2006)	287	248	103	54
# Wrong (2003–2006)	17,326	14,102	5,694	3,013

Summary and Future Work

MIRA:

- Produces a highly recall-oriented answer stream,
- Covers more questions than the other answer producers in IRSAW, and
- Returns the largest number of correct answer candidates.

Future work:

- Return additional answer support for temporal deictic expressions
- Support processing list questions

Selected References

Glöckner, Ingo; Sven Hartrumpf; and Johannes Leveling (2007). Logical validation, answer merging and witness selection – a case study in multi-stream question answering. In *Proceedings of RIAO 2007, Large-Scale Semantic Access to Content (Text, Image, Video and Sound)*. Pittsburgh, USA: C.I.D.

Hartrumpf, Sven and Johannes Leveling (2007). Interpretation and normalization of temporal expressions for question answering. In *Evaluation of Multilingual and Multi-modal Information Retrieval: 7th Workshop of the Cross-Language Evaluation Forum, CLEF 2006* (edited by et al., Carol Peters), volume 4730 of *LNCS*, pp. 432–439. Berlin: Springer.

Leveling, Johannes (2006). On the role of information retrieval in the question answering system IRSAW. In *Proceedings of the LWA 2006, Workshop Information Retrieval*, pp. 119–125. Hildesheim, Germany: Universität Hildesheim.