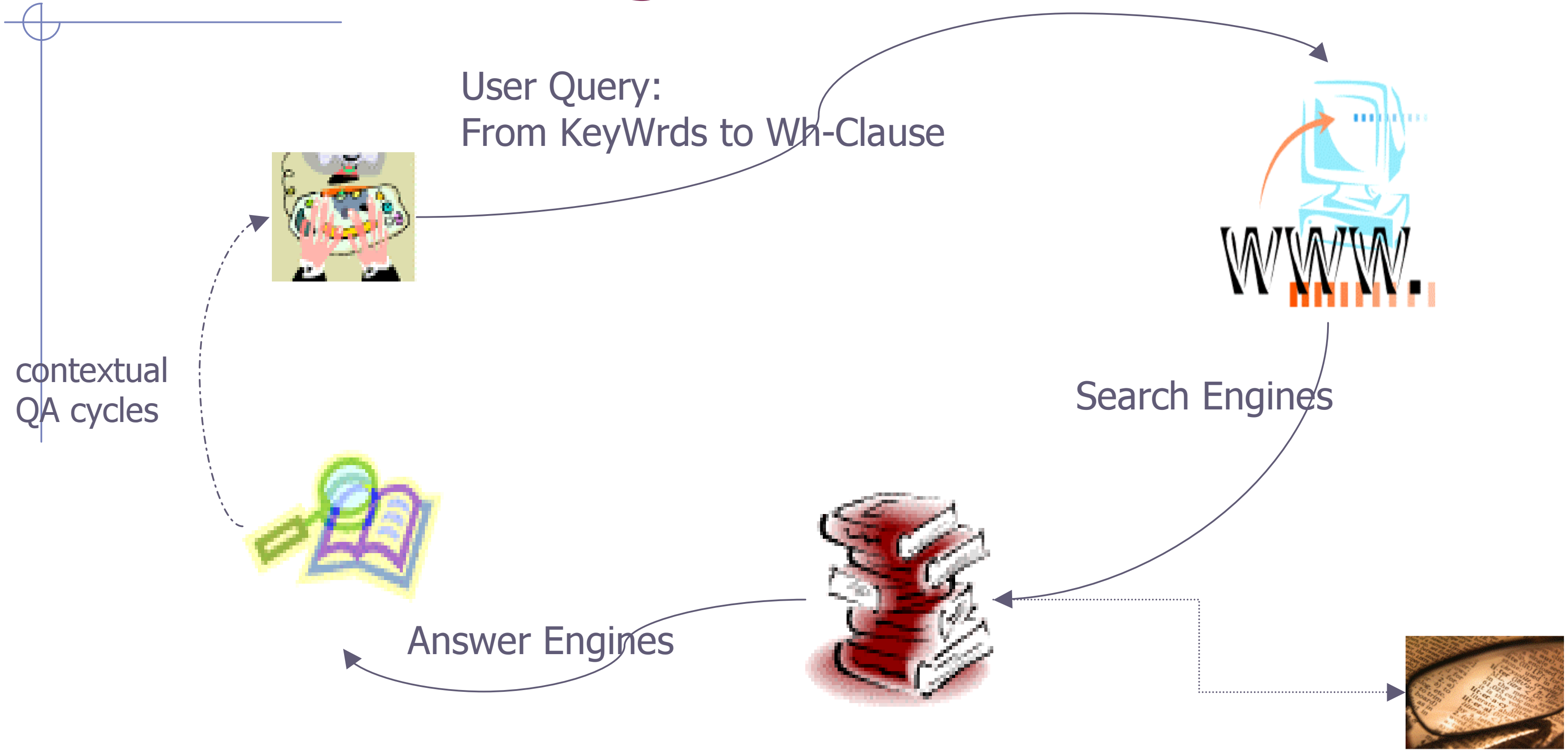
Mining Answers in German Web Pages

Günter Neumann & Feiyu Xu
Language Technology Lab
German Research Center for Artificial Intelligence
DFKI, Saarbrücken

Source: G. Neumann, WI2003, Halifax, CA

Motivation: From Search Engines to Answer Engines



Shift more „interpretation effort" to machines

User stills carries the major efforts in understanding

Source: G. Neuman, WI2003, Halifax, CA

WAG the Web

Web-based answer extraction system for extracting sentence and exact answers from German/(English) web pages.

- Shallow question/answering approach:
  - Named Entity (NE)-recognition and morpho-syntactic analysis as linguistic preprocessing
  - NE-based inverted index for paragraph, sentence and answer selection (NE-based „text zooming“)
  - Token/NE overlap approach for exact answer strings

- Approach can be scaled up for processing different questions types (factoid, list, template questions)

- Evaluation using question-answer pairs extracted from a popular German quiz book.

Source: G. Neuman, WI2003, Halifax, CA

Query Formulation

Who won the Nobel Prize 2000 in chemistry?

will be expressed as:

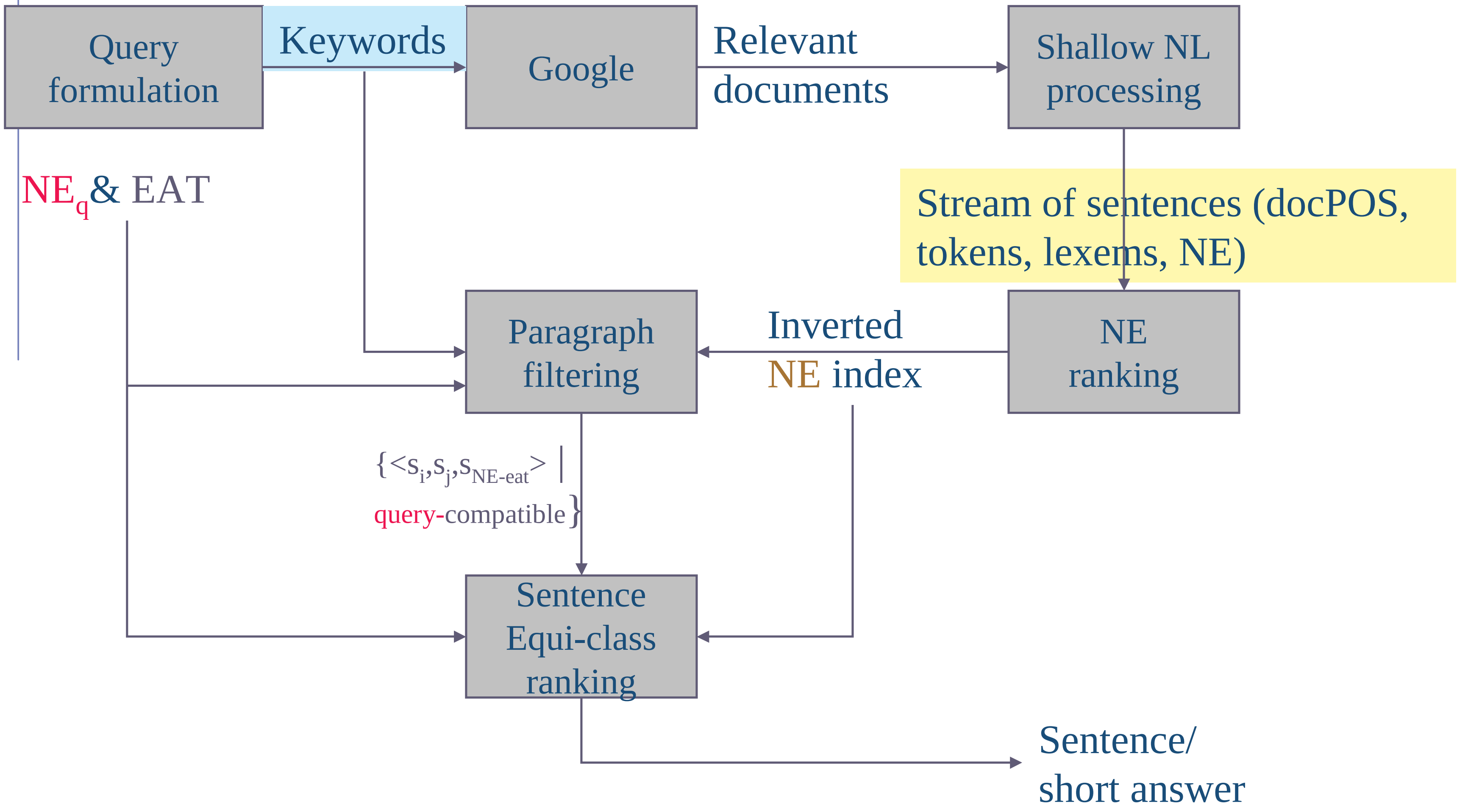
$$\left\{ \begin{bmatrix} T: & tok \\ V: & \text{won} \end{bmatrix}, \begin{bmatrix} T: & tok \\ V: & \text{Nobel Prize} \end{bmatrix}, \begin{bmatrix} T: & DNE \\ V: & 2000 \end{bmatrix}, \begin{bmatrix} T: & tok \\ V: & \text{chemistry} \end{bmatrix}, \begin{bmatrix} T: & PNE \\ V: & ? \end{bmatrix} \right\}$$

Expected answer type (EAT) as NE-type with value ?

Also possible:
multiple EATs (for template-based questions)

Source: G. Neuman, WI2003, Halifax, CA

WAG‘s Control Flow



Source: G. Neuman, WI2003, Halifax, CA

NE Ranking

- Search for *all possible* NEs in the N-Google-retrieved documents D_N covered by the NE grammars of our shallow NLP system
- Note: our NE-recognizer also performs NE-reference resolution (e.g., Martin Marietta Corp. = Marietta)

NE-weighting schema

- Each NE gets a weight according to
  - Absolute frequency of NE in D_N
  - Document frequency of NE
  - Document ranking determined by Google

- NEs mentioned very often in different documents (ranked high by Google) receive high rank

→ Answer extraction will be effected by redundancy and multiple documents

Source: G. Neuman, WI2003, Halifax, CA

Paragraph Selection

Note: each document is a list of sentences (a sentence is a sequence of tokens and NE's as identified by our NLP system)

- Step 1:
  Compute an inverted NE index list (subclassified by types)
  - Each found NE has a pointer into all sentences in D_N in which it occurs
- Step 2:
  For each NE which is type-compatible with EAT
  - Extract a set of paragraph candidates $\{s_1 s_2 s_{NE}, s_2 s_{NE} s_3, s_{NE} s_3 s_4\}$
  - Select the one with highest number of matching query-keywords (+ their distance from NE)

Robustness:

For relevant sentences: no need to contain NE_{EAT} (but a referring general NP); Hence, can tolerate NE-gaps caused by the NE-recognizer, and helps to guide NP-reference resolution

Source: G. Neuman, WI2003, Halifax, CA

Sentence Class Ranking

Collect all sentences from all selected paragraphs with same rank into a class

$$EC = (olToken + olNE) * (1.5 + \frac{\sum_{i=1}^{n} r_{NE(EAT)_i}}{n})$$

Ranking:

number of overlapping elements (tokens/NE_{NOEAT}) between query and sentence;

weighted rank of occurencies of instances of NE_{EAT}

Candidates:

Select from N-best Classes, a sentence and an instances of NE_{EAT}

Source: G. Neuman, WI2003, Halifax, CA

Example Question/Answer Pair

Welches Pseudonym nahm Norma Jean Baker an?
Which pseudonym did Norma Jean Baker use?

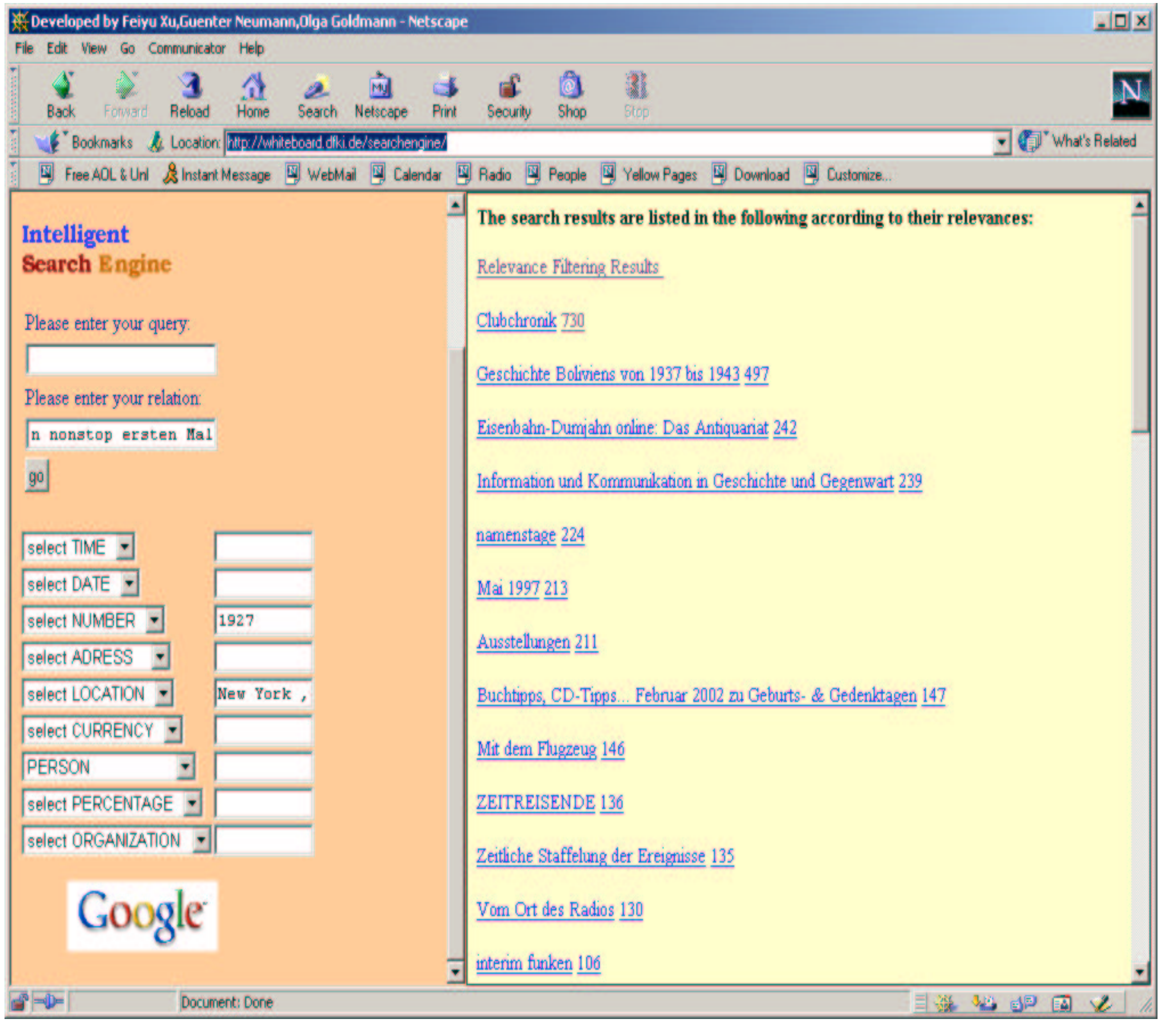
<eclass rank='7.254032855348923'>
<sentence url='http://www.beatlesfan.de/alle.htm'>
Marilyn Monroe war der Kuenstlername von Norma Jean Mortenson, auch bekannt als Norma Jean Baker
</sentence>
<exact-answer type ='PERSON'>
<name rank='0.6029282321968642'>
Marilyn Monroe
</name>
<name rank='0.024088195477597402'>
Norma Jean Mortenson
</name>
</exact-answer></eclass>

Source: G. Neuman, WI2003, Halifax, CA

Query and Google Retrieval

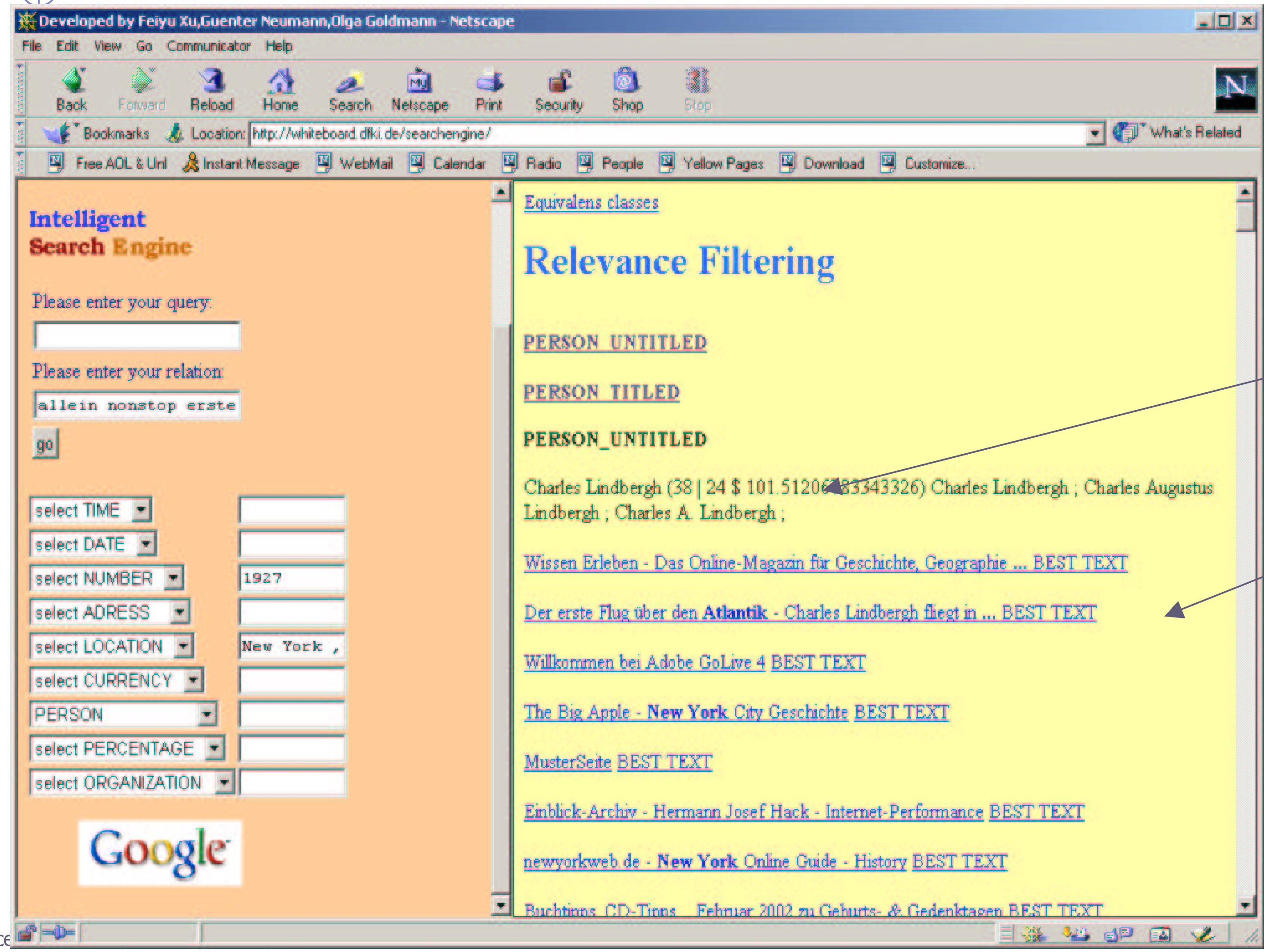
Wer flog 1927 zum ersten Mal allein und nonstop von New York nach Paris über den Atlantik?

Who flew for the first time alone and nonstop from New York to Paris over the Atlantic in 1927?



Source: G. Neuman, WI2003, Halifax, CA

Inverted list of instances of expected answer type

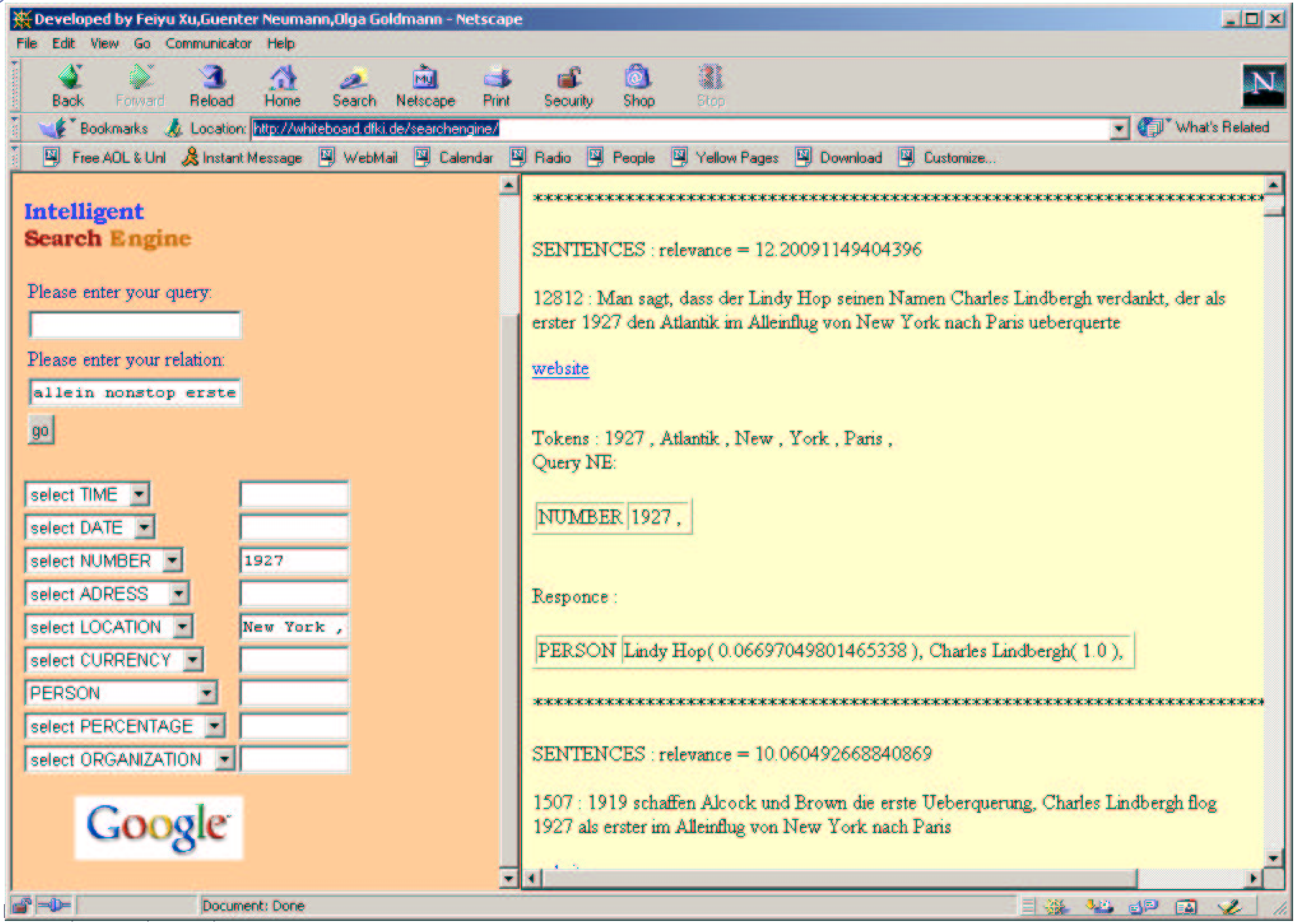


Best NE_{EAT}
38 times found in 24 different documents (incl. NE-references)

Referring documents

Source

Best sentence and exact answer



Source: G.

Initial Experiments

- Corpus:
  - Query/answer pairs from popular German quize book
    - Questions from more than 40 subject areas: philosophy, nature science, history, geography, culture, sport, etc.
    - Questions formalized independ of any QA system in mind
  - Two test environments: restricted vs. extended
- Comparision of WAG results with M=5 top Google-snippets
- Measure:
  - Recall: the percentage of the questions answered correctly compared to all questions of the test corpus
  - Mean Reciprocal Rank (MRR) for N questions and reciprocal rank of found answer for quest I (1/1, ½, …, 1/5 or 0)

$$MRR = \frac{\sum_{i=1}^{N} \frac{1}{rank_i}}{N}$$

Source: G. Neuman, WI2003, Halifax, CA

Two Test Environments

- Restricted
  - 20 person-related and 19 location-related
  - Assume: NE-instances of expected answer types are covered by our NE-recognizer
    - *Who became Israel's first prime minister in 1949?*
    - *In which former soviet state was the nuclear power plant Chernobyl?*
- Extended
  - 25 location-related, 8 organization-related, 35 person-related
  - the expected answer type of a question is not always covered by our NE-recognizer
    - *What's the name of the organisation found in 1971, which helps people all over the world?*
    - *Which boy group did the English singer Robbie Williams belong to before his Solo Career?*

Source: G. Neuman, WI2003, Halifax, CA

Average Results of Restricted Test

Person

Location

Average performance of the Google snippets

Person

Google snippet	MRR (N=5)	correct top1	correct top3
all questions	0.103	0.3	0.35
Excluding no document cases	0.122	0.35	0.41

Location

Google snippet	MRR (N=5)	correct top1	correct top3
all questions	0.092	0.21	0.52

Average performance of WAG exact answers

Person

WAG exact answer	MRR (N=5)	correct top1	correct top3
all questions	0.212	0.45	0.55
excluding no document cases	0.236	0.53	0.64

Location

WAG exact answer	MRR (N=5)	correct top1	correct top3
all questions	0,135	0.31	0.42

Average performance of WAG sentence answers

Person

WAG sentence	MRR (N=5)	Recall top1	Recall top3
All questions	0,216	0.5	0.55
excluding zero document cases	0,254	0.59	0.64

Location

WAG sentence	MRR (N=5)	Recall top1	Recall top3
all questions	0,126	0.26	0.37

Source: G

Average Results of Extended Test

Person

Location

Average performance of the Google snippets

Person:

Google snippet	MRR (N=5)	Recall Top1	Recall top3
all questions	0,126	0.34	0.48

Location:

Google snippet	MRR (N=5)	Recall top1	Recall top3
all questions	0.113	0.216	0.6

Average performance of WAG exact answers

Person:

WAG exact answer	MRR (N=5)	Recall top1	Recall Top3
all questions	0,127	0.314	0.4

Location:

WAG exact answer	MRR (N=5)	Recall top1	Recall top3
all questions	0.16	0.44	0.56

Average performance of WAG sentence answers

Person:

WAG sentence	MRR (N=5)	Recall top1	Recall Top3
all questions	0,154	0.4	0.485

Location:

WAG Sentence	MRR (N=5)	Recall top1	Recall top3
all questions	0,183	0.32	0.72

Source: G

Summary of Evaluation

- Google in Restricted (39)

Metric	snippet
MRR (N=5)	0.0975
Recall top1	0.255
Recall top3	0.435

- Google in Extended (68)

Metric	snippet
MRR (N=5)	0.136
Recall top1	0.310
Recall top3	0.61

- WAG in Restricted

Metric	exact answer	Sentence
MRR (N=5)	0.174	0.171
Recall top1	0.38	0.38
Recall top3	0.48	0.46

- WAG in Extended

Metric	exact answer	Sentence
MRR (N=5)	0.126	0.166
Recall top1	0.29	0.365
Recall top3	0.445	0.61

Source: G

Qualitative Analysis

- Does the Quiz Book provide actually valid answer keys?
  - Tradeoff between WAG answers and Quiz book answer keys
  - Incompleteness of the provided answers
- Exact Answer and Sentence Answer
  - Sentence answer is more robust than exact answer
  - Exact answers provide more clear response in case of non-"well-formed" sentences
- Redundancy and Popular Information
  - Majority polling alone is not sufficient to guarantee the correct answer
  - Most popular information will be ranked higher than the correct answer in some cases
- Anaphora Resolution
  - Sentences maximally matched with queries contain sometimes anaphoric expressions instead of answers, therefore,
  - Paragraphs can be used as discourse context for anaphor resolution

Source: G. Neuman, WI2003, Halifax, CA

Summary: the lesson we learned

- Web-based answer extraction benefits from robust and efficient LT-technology
- The better NE-recognition is, the better the answer extraction process is (wrt. coverage and fine-grainedness)
- Using a NE-anchored paragraph, sentence and phrase selection strategy helps in
  - Handling multiple documents
  - Taking into account the redundancy on the Web
  - Processing complex (list, template) questions

Source: G. Neuman, WI2003, Halifax, CA

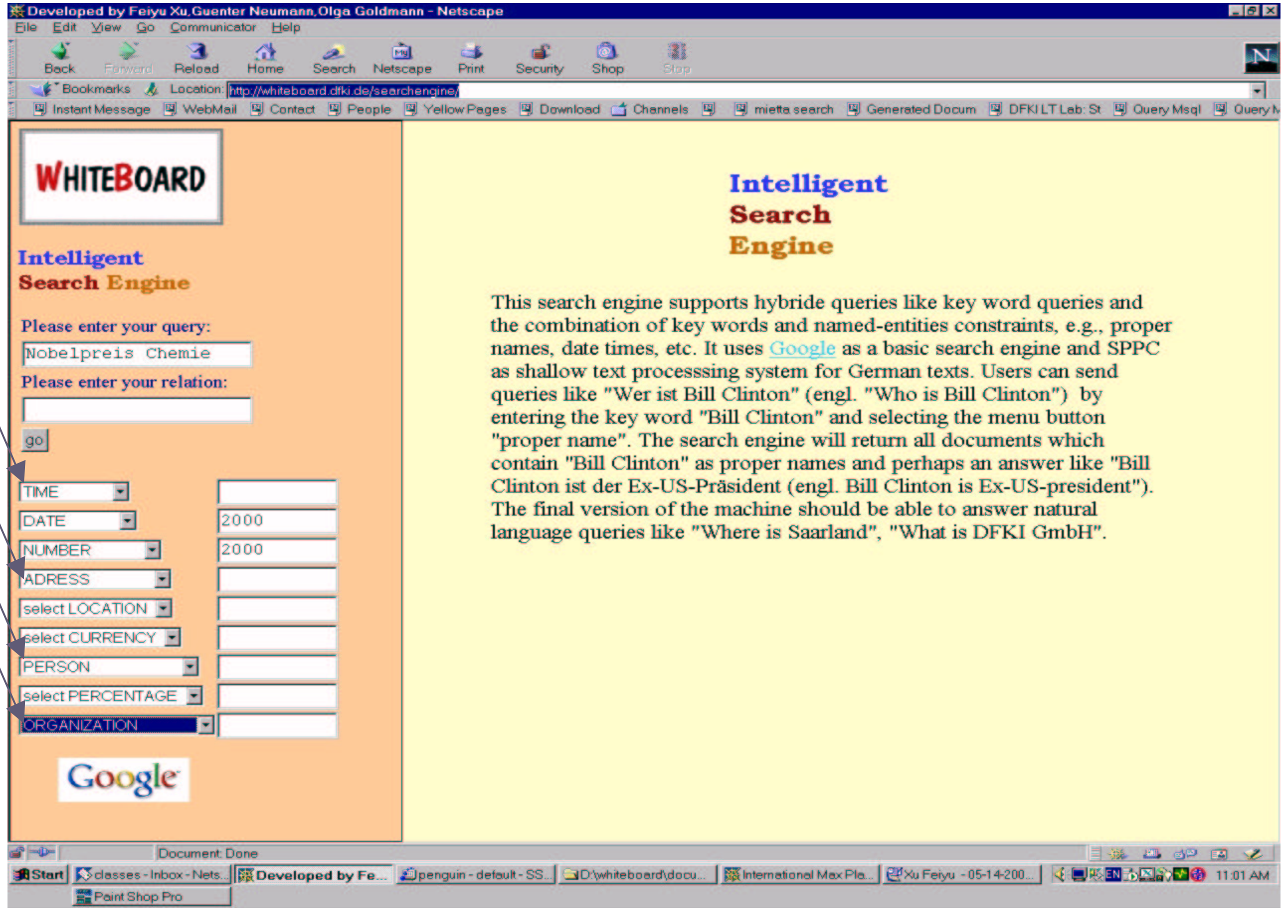
Future work: Cross-language Template-based QA

- Cross-language web-based QA
  - German queries, English documents (and vice versa)
  - First prototype for Trec-like corpus exists (NeumannSacaleanu, Clef-2003)

- Web-based Dynamic Information Extraction
  - Specification of multi-slot questions
  - Extraction/Merging of (partial) templates
    - Single instances
    - Multiple instances (list-based template merging)
  - First prototype, not yet evaluated

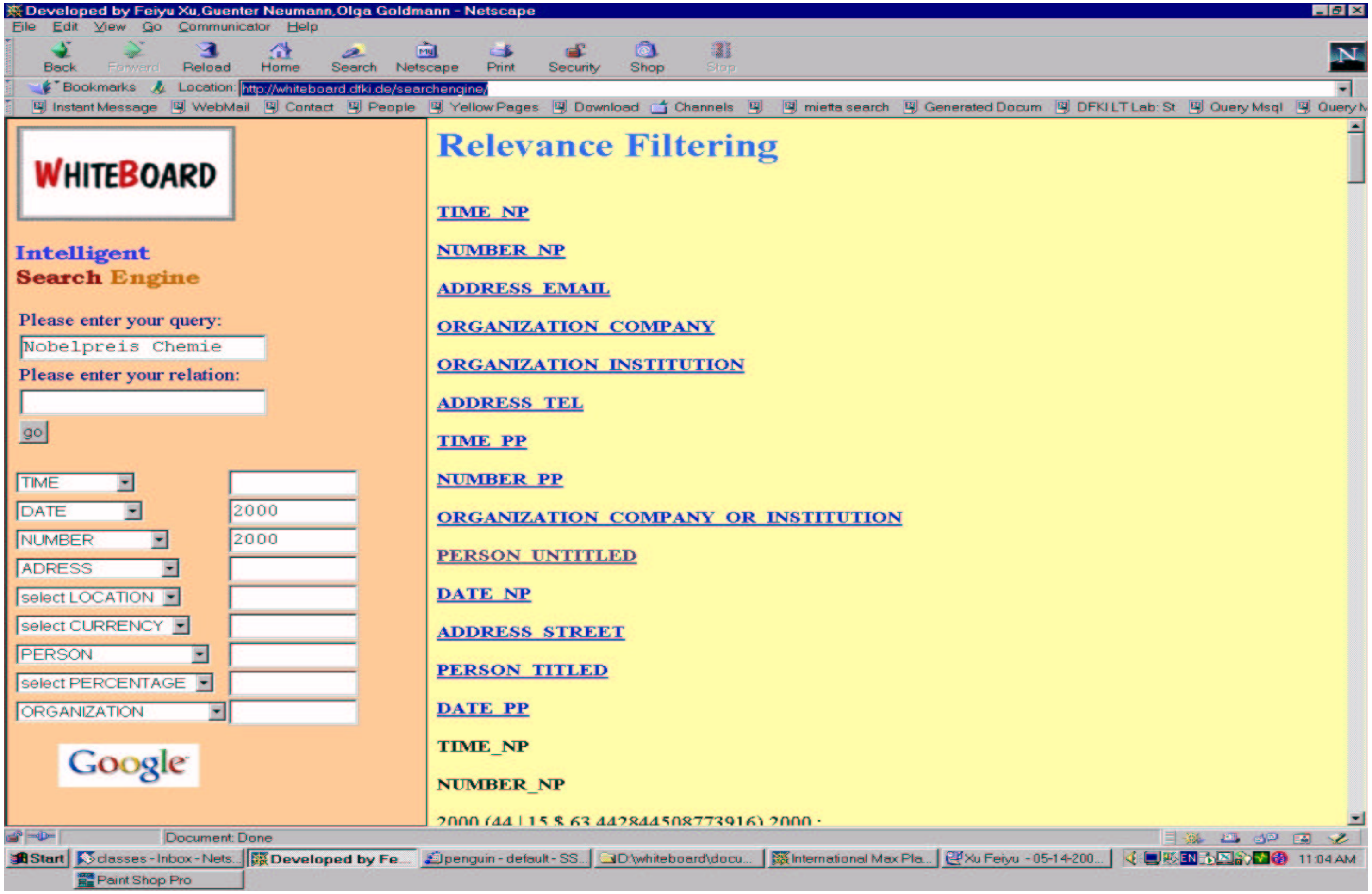
Source: G. Neuman, WI2003, Halifax, CA

Query-Specification: Multiple Expected Answer Types



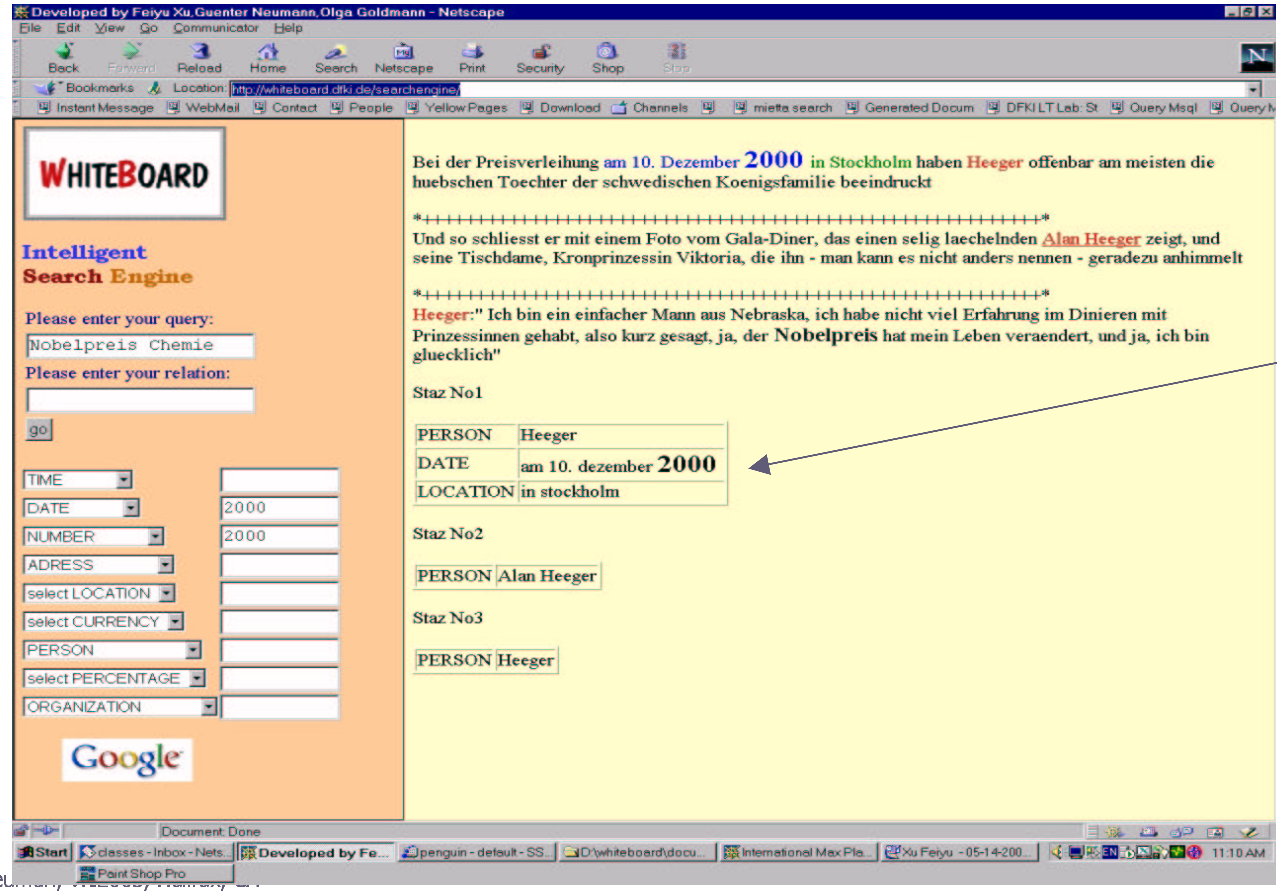
Source: G. Neuman, WI2003, Halifax, CA

Ranking according to all processed NE types



Source: G. Neuman, WI2003, Halifax, CA

Template extraction through merging



So far: Do handle Single Instance, in single sentence

Next step: merging of partial templates

Source: G. Neu

Done –

Thank you for your attention!

Source: G. Neuman, WI2003, Halifax, CA

WAG Example

- Paragraph retrieval for template extraction/template merging/template mining
  - <sentence>

    <DATE>10.2000 </DATE> ChemieChemie-Nobelpreis 2000: Elektrisch leitende Kunststoffe koennten die Technik des 21. Jahrhunderts bestimmen Durch die Entdeckung leitfaehiger Kunststoffe haben die beiden <Nationality>US-Amerikaner</Nationality> <PN>Alan Heeger </PN> von der <Organization> Universitaet von Kalifornien </Organization> in <Location>Santa Barbara </Location> und <PN> Alan MacDiarmid </PN>von der <Organization> Universitaet von Pennsylvania </Organization> und der <Nationality> Japaner </Nationality> <PN> Hideki Shirakawa </PN> von der <Organization> Universitaet Tsukuba </Organization> Anwendungen wie Leuchtdioden oder Anzeigendisplays in Mobiltelefonen moeglich gemacht.

    </sentence>

[PersonName: Alan Heeger
Organization: Universitaet von Kalifornien
Location: Santa Barbara
Nationality: US-Amerikaner]

[PersonName Hideki Shirakawa
Organization Universitaet Tsukuba
Location Santa Barbara
Nationality Japaner]

[PersonName Alan MacDiarmid
Organization Universitaet von Pennsylvania
Location Santa Barbara
Nationality US-Amerikaner]

Source: G. Neuman, WI2003, Halifax, CA