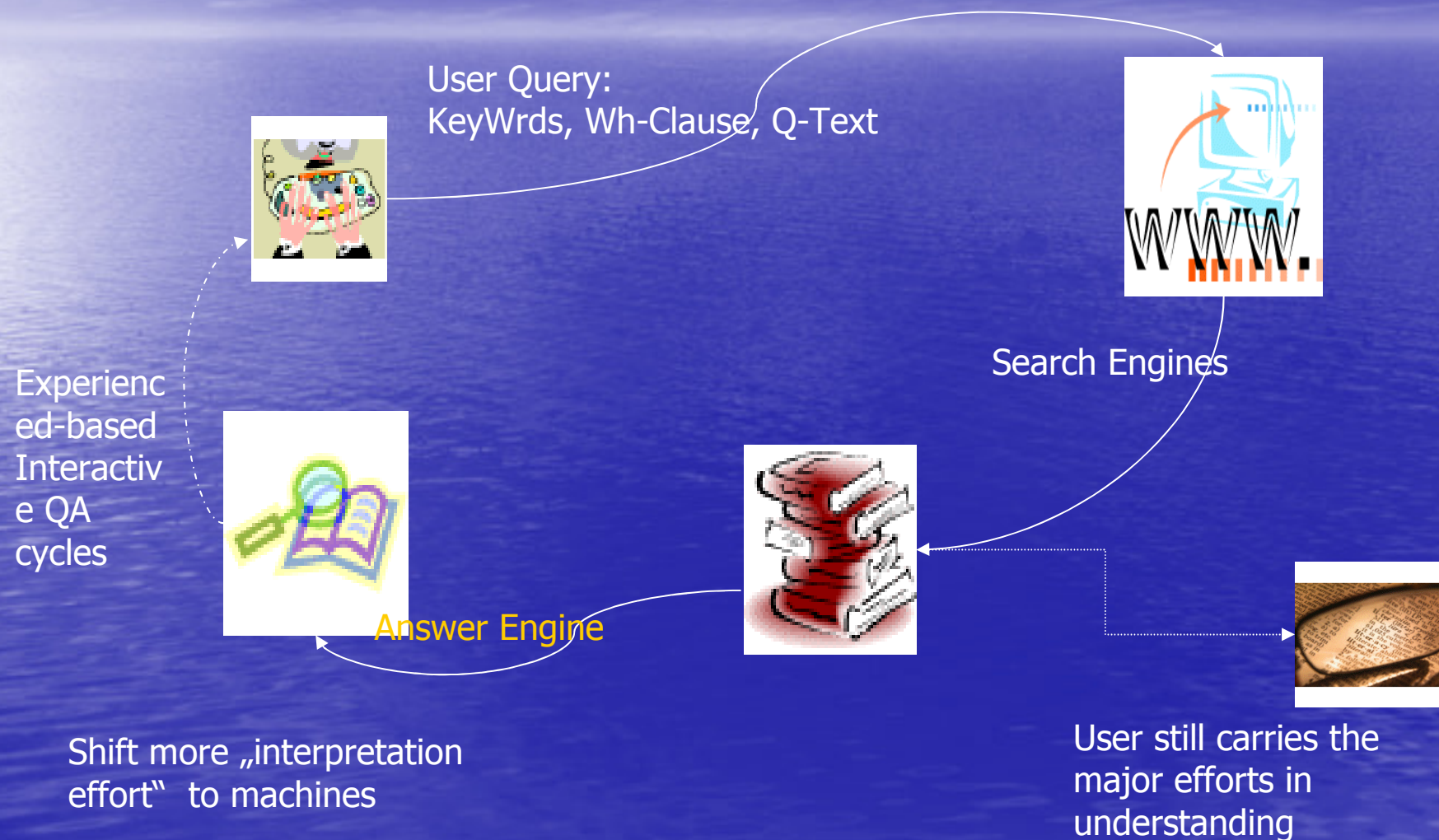# Language Independent Answer Prediction from the Web
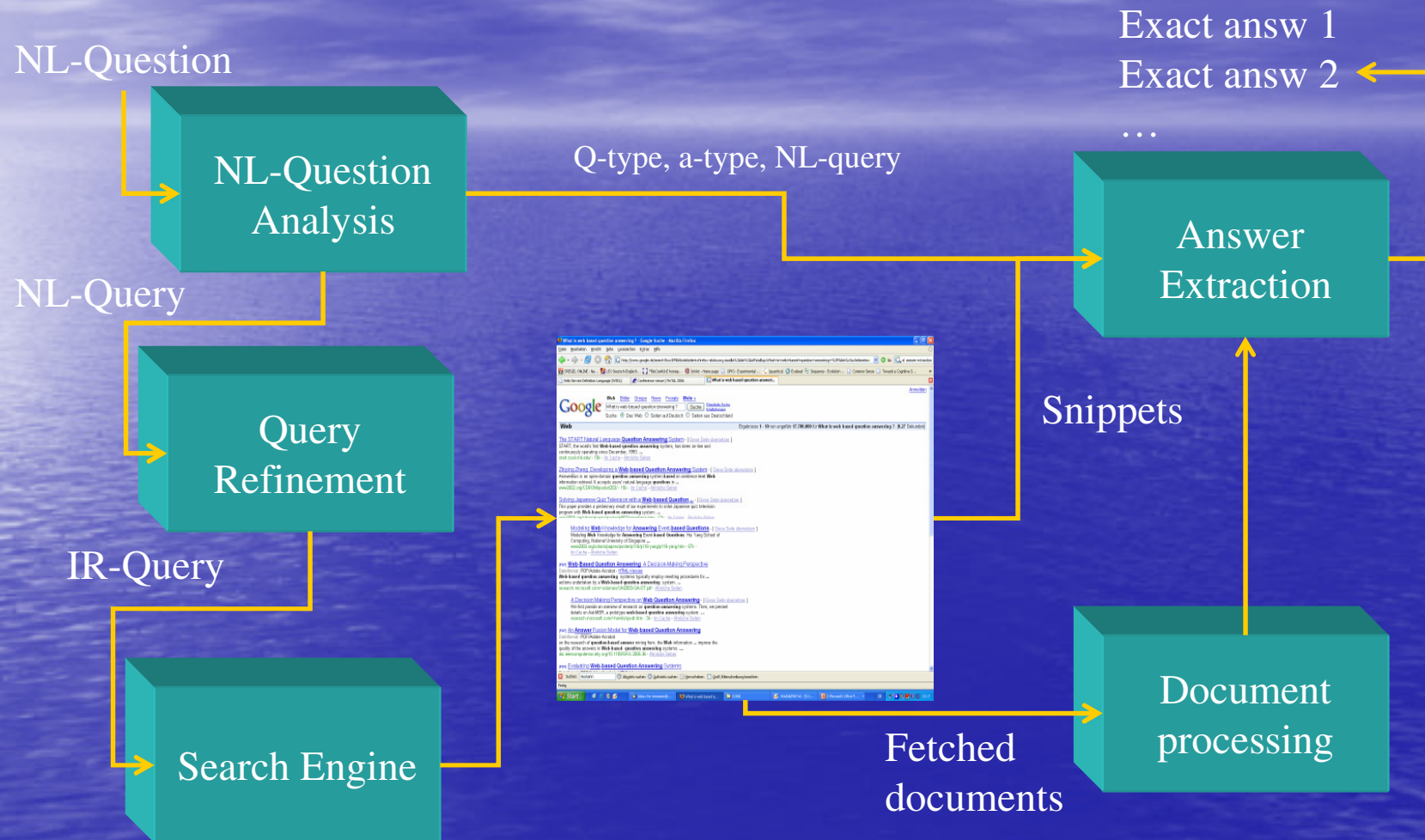
Alejandro Figueroa & Günter Neumann

Language Technology Laboratory
German Research Center for Artificial Intelligence
DFKI, Saarbrücken

Motivation: From Search Engines to Answer Engines

User Query:
KeyWrds, Wh-Clause, Q-Text

Search Engines

Experienced-based Interactive QA cycles

Answer Engine

Shift more „interpretation effort" to machines

User still carries the major efforts in understanding

Web-based Question Answering: SOA

NL-Question

NL-Question Analysis

Q-type, a-type, NL-query

Exact answ 1
Exact answ 2
…

Answer Extraction

NL-Query

Query Refinement

Snippets

IR-Query

Search Engine
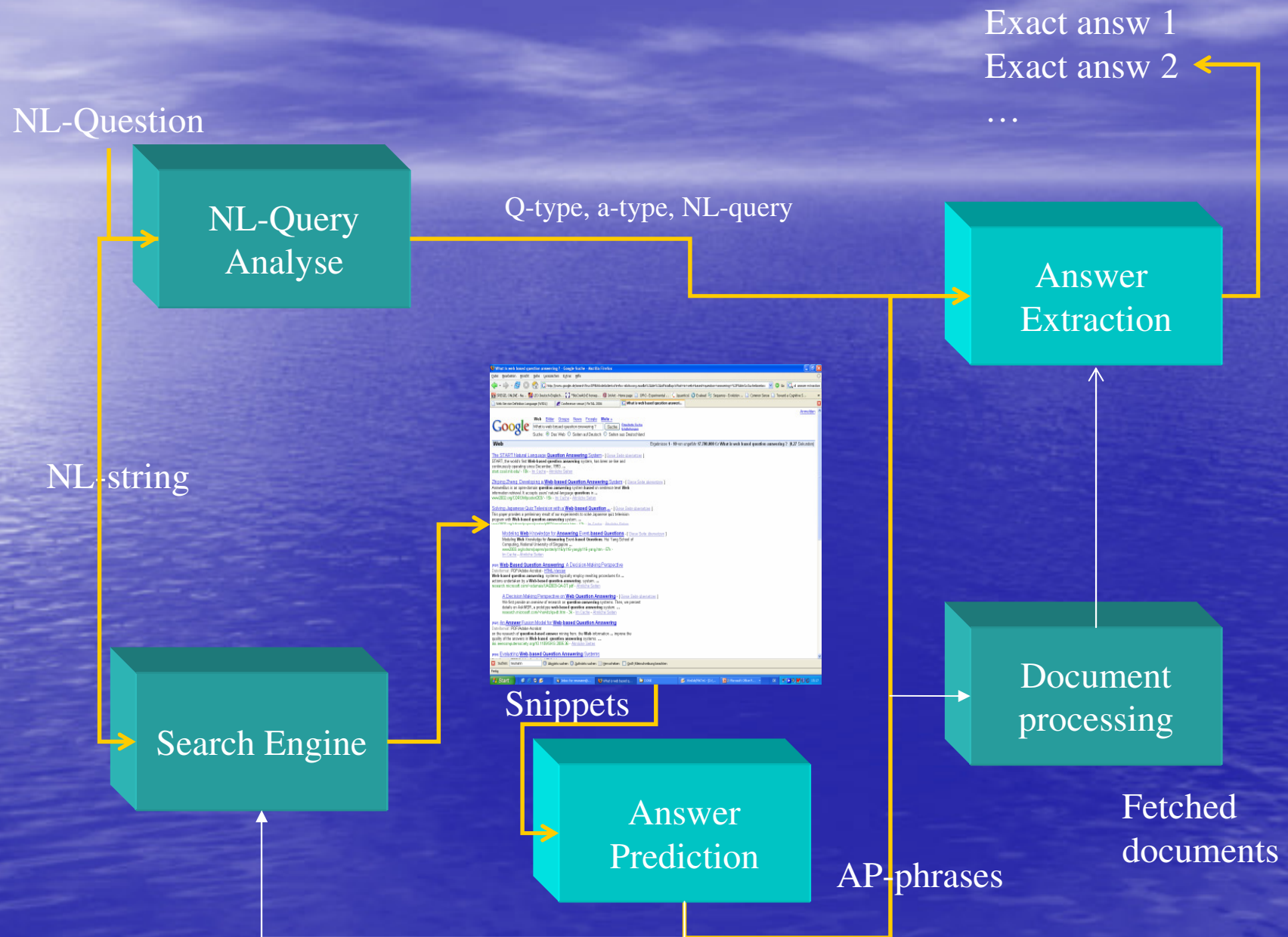
Document processing

Fetched documents

Problematic:
NL-question-driven Query Refinement
Usually large set of snippets/documents have to be consulted

Web-based Question Answering: Our Approach

Exact answ 1

Exact answ 2

…

NL-Question

NL-Query Analyse

Q-type, a-type, NL-query

Answer Extraction

NL-string

Snippets

Search Engine

Answer Prediction

AP-phrases

Document processing

Fetched documents

Answer prediction ...

- ... identifies and extracts those substrings (called the AP-phrases) from a set of snippets/text fragments
  - Which are either paraphrases of substrings of the NL question, or
  - Which contain exact answer strings
  - E.g., „Who is the primer minister of Great Britain ?"
    - Possible AP-phrases: „United Kingdom", „Tony Blair"

- Notice:
  - Often, the first N-snippets do not contain the answer and even point to answerless documents (we cannot rely on the first 3-5 snippets/documents; ~ low MRR of snippets)

  - We need to consider several snippets/documents in order to make use of redundancy (only checking the first is not enough even if it contains the answer)

  - Snippets are usually not very much linguistically well-formed
  - They are computed online with very fast, but cheap methods

Web **Bilder** Gro

Google

where is the tal conference series taking pla| | Suche | Erweiterte Suche
Einstellungen

Suche: ⊙ Das Web ◯ Seiten auf Deutsch ◯ Seiten aus Deutschland

Web   Ergebnisse **1 - 10** von

Meinten Sie: where is the *tall* conference series taking place in 2006?

President Discusses Democracy in Iraq with Freedom House - [Diese Seite übersetzen]
This month I've given a **series** of speeches on recent events in Iraq and how ... As Peter
mentioned, there is a philosophical debate **taking place** in the ...
www.whitehouse.gov/news/releases/**2006**/03/**2006**0329-6.html - 102k -
Im Cache - Ähnliche Seiten

President Discusses War on Terror, Progress in Iraq in West Virginia - [Diese Seite übersetzen]
And there was some awful violence, some reprisals **taking place**. ... I spoke in Cleveland,
gave a press **conference** yesterday -- spoke in Cleveland Monday, ...
www.whitehouse.gov/news/releases/**2006**/03/**2006**0322-3.html - 88k -
Im Cache - Ähnliche Seiten
[Weitere Ergebnisse von www.whitehouse.gov]

Konferenser - Conferences - [Diese Seite übersetzen]
The main purpose of the **TAL conference series** is to bring together ... Do we fully understand
the changes which are **taking place** and appreciate the threats ...
www.tolk.su.se/0607-12.html - 204k - Im Cache - Ähnliche Seiten

Expert System - News - [Diese Seite übersetzen]
"**TAL Conference 2006**" at the Ministry of Communications ... Expert System at the workshop
taking place in Bologna within the eCitizen European project ...
www.expertsystem.it/vetrinanews.asp?lang=1&id=1536 - 22k - Im Cache - Ähnliche Seiten

Informed Comment - [Diese Seite übersetzen]
"I was impressed by him," said Bush during a press **conference**. ... to attack Turkmen in **Tal**
Afar is also not wise from the point of view of ethnic politics. ...

E - Na... | LEO Deutsch-Englisch... | *BioCreAtIvE homep... | IntAct - Home page | DFKI - Experimental ... | /quantico/ | Exalead | Sequence - Evolution ... | Common Ser

conference series taking pla... | DFKI - Experimental Question Answer...

DFKI - MULTI LINGUAL WEB QUESTION ANSWERING SYSTEI

Enter Your Question: where is the tal conference series taking place in 2006

WHERE IS THE TAL CONFERENCE SERIES TAKING PLACE

[11784] AFAR

1. [18.6748] FOLLOWING AN EXPLOSION IN THE CITY OF *AFAR* ON 9 MAY MANY CIVILIAN IRAQI MINORITIES TOOK PLACE 27 MAY
   - FOLLOWING AN EXPLOSION IN THE CITY OF AFAR ON 9 MAY MANY CIVILIAN IRAQI MINORITIES TOOK PLACE 27 MAY 2006 . THE CONFEREN
   - http://www.uniraq.org/documents/hr%20report%20may%20jun%202006%20en.pdf

2. [13.8904] TO ATTACK TURKMEN IN *AFAR* IS ALSO NOT WISE FROM THE POINT OF VIEW OF ETHNIC POLITICS ...
3. [13.7759] I TALKED ABOUT A CITY NAMED *AFAR* THE OTHER DAY IN A SPEECH I GAVE IN CLEVELAND ...
4. [12.0858] IM GOING TO TELL YOU THE STORY OF A NORTHERN IRAQI CITY CALLED *AFAR* ...

[57] MAIN

1. [26.0614] THE SYMPOSIUM WILL TAKE PLACE IN *TURKU IN THE MAIN PURPOSE OF THE TAL* CONFERENCE SERIES IS TO BRIN(

[9.1314] TURKU

1. [26.0614] THE SYMPOSIUM WILL TAKE PLACE IN *TURKU IN THE MAIN PURPOSE OF THE TAL* CONFERENCE SERIES IS TO BRIN(

OTHER RESULTS

[33.7024] A CALL FOR PAPERS IS TAKING PLACE FOR THE *9TH INTERNATIONAL CONFERENCE* ....

[31.661] TAL CONFERENCE 2006 AT THE MINISTRY OF COMMUNICATIONS EXPERT SYSTEM AT THE WORKSHOP TAKING PLACE IN BOLOGNA WITHIN TH

[30.8202] WITH THE AND SORTED THROUGH A IN WHATS TAKING PLACE IN THEIR CONTINENT AND HAVE BEEN ATTENTIVE ....

[30.6709] THE *MAIN PURPOSE OF THE TAL* CONFERENCE SERIES IS TO BRING TOGETHER DO WE FULLY UNDERSTAND THE CHANGES WHICH ARE TAKIN

[30.2252] THERE IS A PHILOSOPHICAL DEBATE TAKING PLACE IN THE ....

[29.5784] 15TH APRIL 2006 THE 10TH IINTERNATIONAL CONFERENCE OF ISSEI TAKING PLACE IN MALTA ....

[29.4403] DOUBLETREE HOTEL AT *WARREN PLACE IN CHANGES* TAKING PLACE IN TODAY?S GLOBAL ....

Technological Roadmap for this work

Data-driven
No initial query expansion/
refinement without
initial documents

**Language independent
answer prediction**
- No NLP components
- No language model

**Unsupervised
data-management**
- No parameter smoothing
- No restrictions on
  length of AP-phrases
- No fixed window size

How far can we go with this ?

Answer Prediction as text zooming: core steps

N-snippets +
NL-question

Document Construction:
Token set W
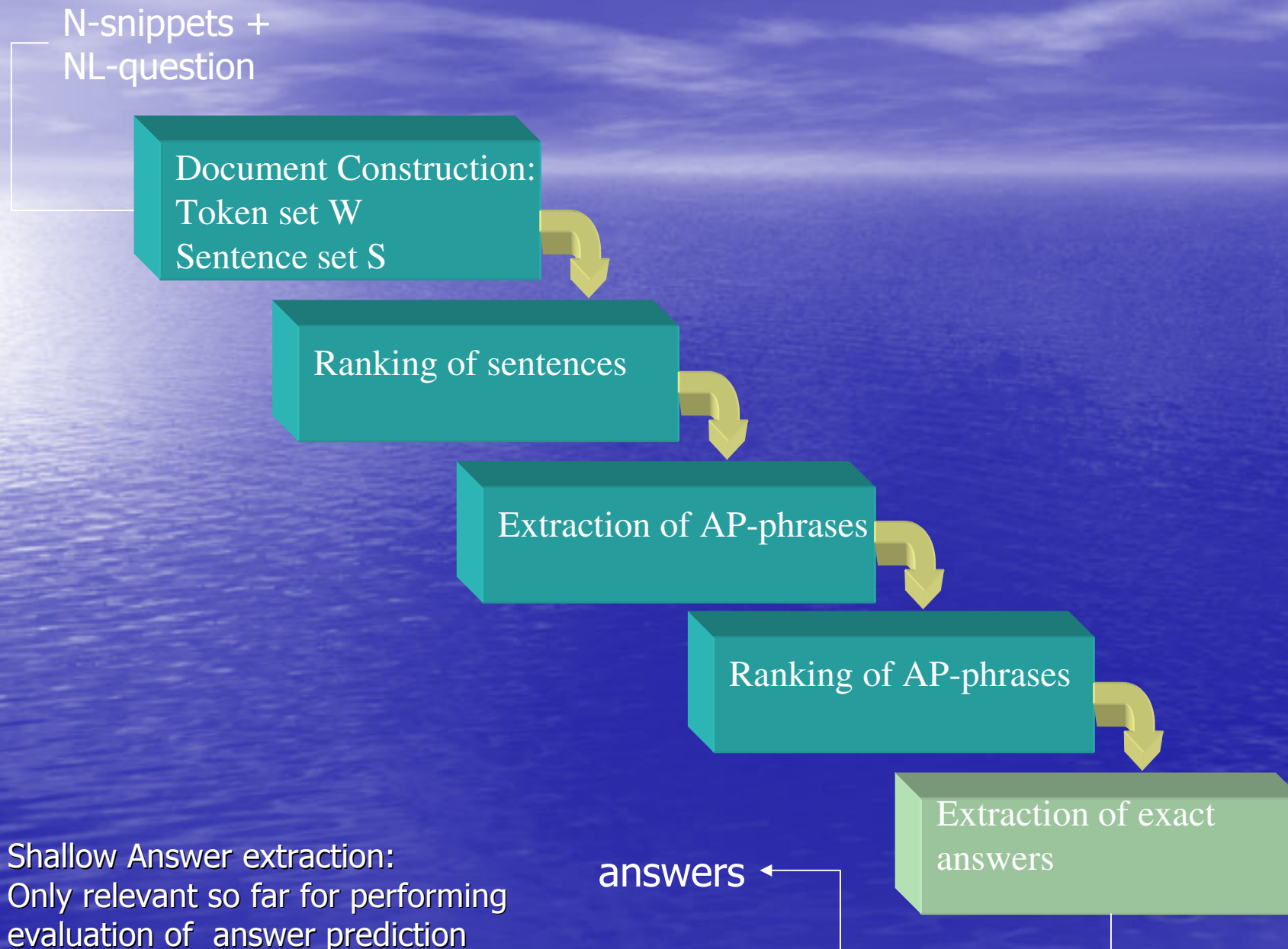Sentence set S

Ranking of sentences

Extraction of AP-phrases

Ranking of AP-phrases

Extraction of exact answers

answers

Shallow Answer extraction:
Only relevant so far for performing
evaluation of answer prediction

Document Construction/Representation

- Sentences from snippets & NL question
  - Only very simple/local tokenizer & sentence markers

- Collect global statistics

$$X_{sik} = \begin{cases} 1 & \textit{if the word } w_i \textit{ is in } S_s \textit{ at position } k. \\ 0 & \textit{Otherwise.} \end{cases}$$

  - Word pair-distance frequency (respecting order)

$$freq(\omega_i, \omega_j, \varepsilon) = \sum_{s=1}^{\sigma} \sum_{k=\varepsilon+1}^{len(S_s)} X_{si(k-\varepsilon)} X_{sjk}$$

- A document D is represented by the following elements:

$$D = \{ <\omega_i, \omega_j, \varepsilon, freq(\omega_i, \omega_j, \varepsilon)>, \\ \forall i, j, \varepsilon, \\ freq(\omega_i, \omega_j, \varepsilon) > 0, \\ 0 \leq \varepsilon \leq Y \}$$

Length of the longest Sentence in D

All possible units |W|x|W|xY, |W| relative small (N-snippets)

Ranking of Sentences

- A matrix for each sentence:

$$M_{ij}(S_s) = \begin{cases} freq(\omega_i, \omega_j, \varepsilon) & if \ i < j \\ freq(\omega_j, \omega_i, \varepsilon) & if \ j > i \\ 0 & otherwise \end{cases}$$

- Filtering: ignore low frequent elements ($\zeta=2$ globally).

$$\forall i, j \ M_{ij} \leq \zeta \Rightarrow M_{ij} = 0$$

To avoid bias of long sequences of low correlated words

- *rank(Ss)* is given by the *maximal eigenvalue* of M:

$$rank(Ss) = \lambda_{\max}(M(S_s))$$

This eigenvalue gives the amount of „*syntactic bonding force*" captured by the eigenvector related to $\lambda_{\max}$.

Ranking Sentences - Remarks

- Retrieval of snippets is biased by terms in the query
- Snippets not only consists of query terms, they also consists of enriched contextual information
- Our ranking schema identifies strong syntactic patterns in snippets using them to rank the sentences in the snippets
- A high ranked sentence not necessary contains query terms, but might contain the answer

- What is the difference between this approach and a ranking based on *n-grams* (e.g., AskMRS)
  - We do not have any dependency on lengths
  - We do not need to estimate back-off probabilities
  - We do not have the problem that long sentences will tend to have a lower rank than small sentences

Determination of AP-phrases

- Idea: Sequences of pairs of words which occur with a high frequency in M (i.e., in a sentence) are *chains of related words*, that is, our *AP-phrases*.

- Words that do not have a strong relation with any other word in S_s are replaced with a „*" -> defines cutting points for sentences

- Example:
  - „The president of France went on Holidays yesterday"
  - „The president of France * * on Holidays *"
  - „The president of France", „on Holidays"

Ranking of AP-phrases

- For each AP-phrase we combine its bi-gram statistics (global context via snippets) with the rank of its embedding sentence (local context)

$$rank(v) = rank(S_s) * \sum_{b=2}^{\beta} P(B_b \mid B_{b-1})$$

$$P(B_b \mid B_{b-1}) = \frac{\log(freq(B_b - 1, B_{b-1}))}{\log(freq(B_{b-1}))}$$

Log reduces the trend to favor high frequent words.

Note:
an AP-phrase can be mapped to different rank values (if it is extracted from different sentences) -> keep only the highest ranked one.

How to Measure the Quality of an AP-phrase ?

- Remember:
  An AP-phrase is either a paraphrase to a substring of the query or an exact answer string.
  - Basically no NLP components
  - Daten-driven, language independent

- We assume that:
  The distribution of answers in the ranking gives a notion of the potential quality of the answer prediction strategy
  - Use the ranked AP-phrases for extracting exact answers
  - Use simple answer extractors simulating a standard QA

Shallow Answer Extraction

- First Step, determine EAT just be looking up Wh-forms

EAT	Keywords
Date	Wann, When, Cuándo, Qué ano, Welchem Jahr, Que ano
Location	Wo, Where, Dónde, Onde
Person	Wer, Who, Quién, Quem

- Second step, extract terms as exact answer candidates, basically
  - * for query terms and ccw
  - * for numeric characters (who)/ non-numeric (when)
  - Who/when: Answers are terms separated by *
  - Where: terms that match a location name in Wordnet (via Babelfish)

Experiments

- CLEF 2004 corpus
    - QA pairs from 1994/95 newspaper texts
    - When/Who/Where questions for 4 languages

- N=30 Snippets and Google-API

- Two types of answers:
    - **Exact Answer**:
        - Exact matching with the answer provided by CLEF.
    - **Inexact Answer:**
        - Are not exact answers, but they are very close answers:
            - **WHERE**: not only city name, country name is also correct.
            - **WHO**: variants like „G. Bush", „George W. Bush".
            - **WHEN**: „6 1945","1945".

- Inexact answers are important, because we aim for assessing the quality of predicted answers.

Experiments

Results for each question type over all languages.

CA	Total	MRR	NAG(%)	WAG(%)	NAF(%)	1(%)	2(%)	3(%)
WHEN	218	0.60	25.11	10.96	21.46	35.16	5.02	1.8
WHERE	232	0.57	10.77	24.14	20.68	30.60	9.91	3.87
WHO	439	0.38	11.39	27.56	32.57	18.90	6.83	2.73

Distribution of answer candidates (all languages).

CA	NAF(%)	1(%)	2(%)	3(%)
WHEN	33.82	55.42	7.91	2.84
WHERE	31.86	47.00	15.23	5.95
WHO	53.37	30.97	11.19	4.47

Experiments

The results for the individual languages.

CA(EN)	Total	MRR	NAG(%)	WAG(%)	NAF(%)	1(%)	2(%)	3(%)
when	69	0.69	15.69	15.69	17.65	45.10	3.92	1.96
where	64	0.74	7.81	12.5	15.62	53.12	10.93	0
who	148	0.50	7.43	12.83	32.43	33.78	10.14	3.38

CA(DE)	Total	MRR	NAG(%)	WAG(%)	NAF(%)	1(%)	2(%)	3(%)
Wann	58	0.45	36.20	12.07	27.59	22.03	1.17	0
Wo	58	0.46	9.37	18.75	23.43	20.31	12.5	6.25

CA(ES)	Total	MRR	NAG(%)	WAG(%)	NAF(%)	1(%)	2(%)	3(%)
Cuándo	59	0.55	16.64	11.86	23.73	32.20	10.17	11.86
Dónde	63	0.59	10.93	31.25	15.62	26.56	10.93	3.21
Quién	86	0.27	9.65	40.68	28.96	11.72	6.21	2.75

CA(PT)	Total	MRR	NAG(%)	WAG(%)	NAF(%)	1(%)	2(%)	3(%)
Quando	56	0.04	30.76	12.30	42.45	3.08	1.54	0
Onde	47	0.18	10.93	25	20.31	10.93	1.56	4.68
Quem	146	0.14	17.12	29.45	36.30	10.95	4.11	2.05

Experiments - Discussion

- The distribution of answers gives the notion of the quality of the ranking strategy

- Our results do not behave in the same way for all kinds of questions and languages:
  - Question types
  - The shallow nature of the answer extractors
  - The redundancy on the Web
  - Different numbers of mentioning a term

- Lita&Carbonell:2004 report a MRR=0.447 for 296 English temporal questions for exact answer matching

- We conclude that our approach is at least competitive

- Experimental QA web system is online
  - About 5-8 seconds/QA cycle

Future Work

- Query refinement and bootstrapping
  – Exploring user feedback

- Daten-driven approach for answer extraction
  – "Exploring Genetic Algorithms", Master Thesis by Figueroa submitted

- Explore method of answer prediction for other applications, e.g., clustering of sequences and recognition of paraphrases

The End!

Thank you for your attention