



# DFKI at QA@Clef 2007

Günter Neumann, Bogdan Sacaleanu,  
Christian Spurk, Rui Wang

Language Technology Lab at DFKI

Saarbrücken, Germany





☆ DFKI is participating since 2003

- Focus on German monolingual QA and German/English cross-lingual QA
- Promising results so far (acc.): DEDE=43,50%, ENDE=32,98%, DEEN=25.50%

☆ Goal for Clef 2007: increase spectrum of activities

- Consideration of additional language pairs (ESEN, PTDE)
- Participation in QAST pilot task
- Participation in Answer Validation Exercise (AVE)



### ☆ NL question

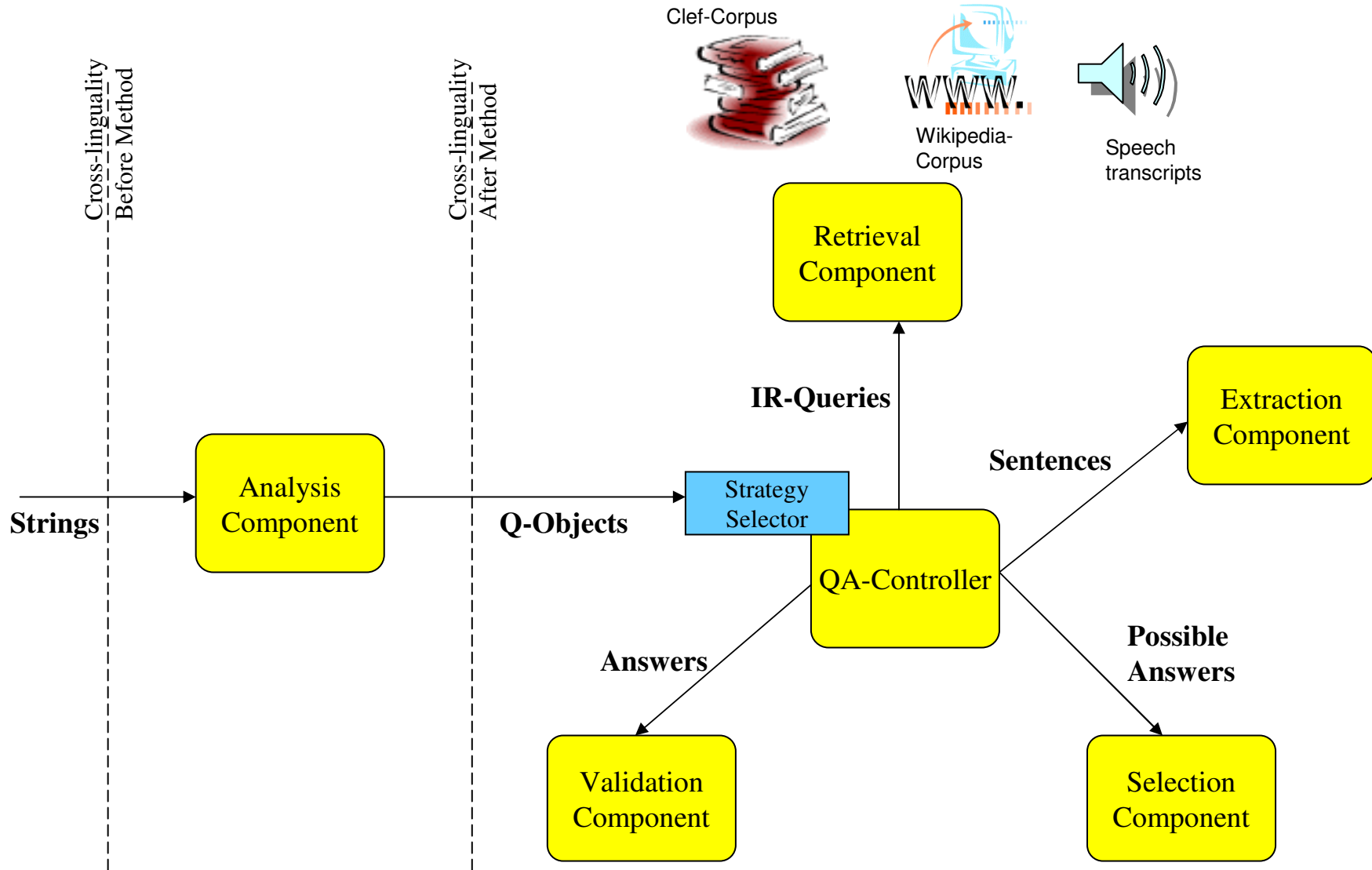
- Declarative description of search strategy and control information
- Analysis should be as complete and accurate as possible
- Use of full parsing and semantic constraints

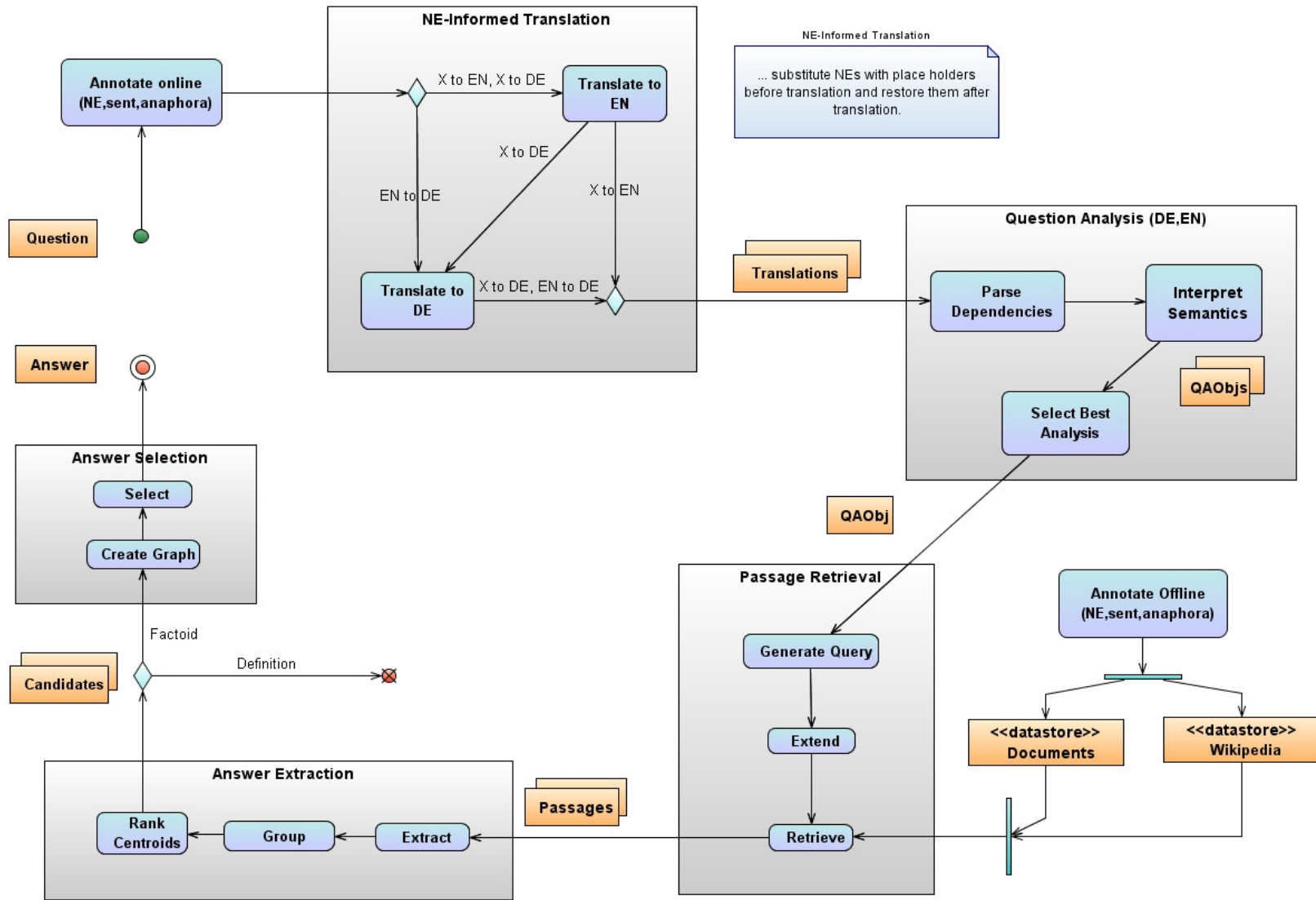
### ☆ Consider document sources as implicit search space

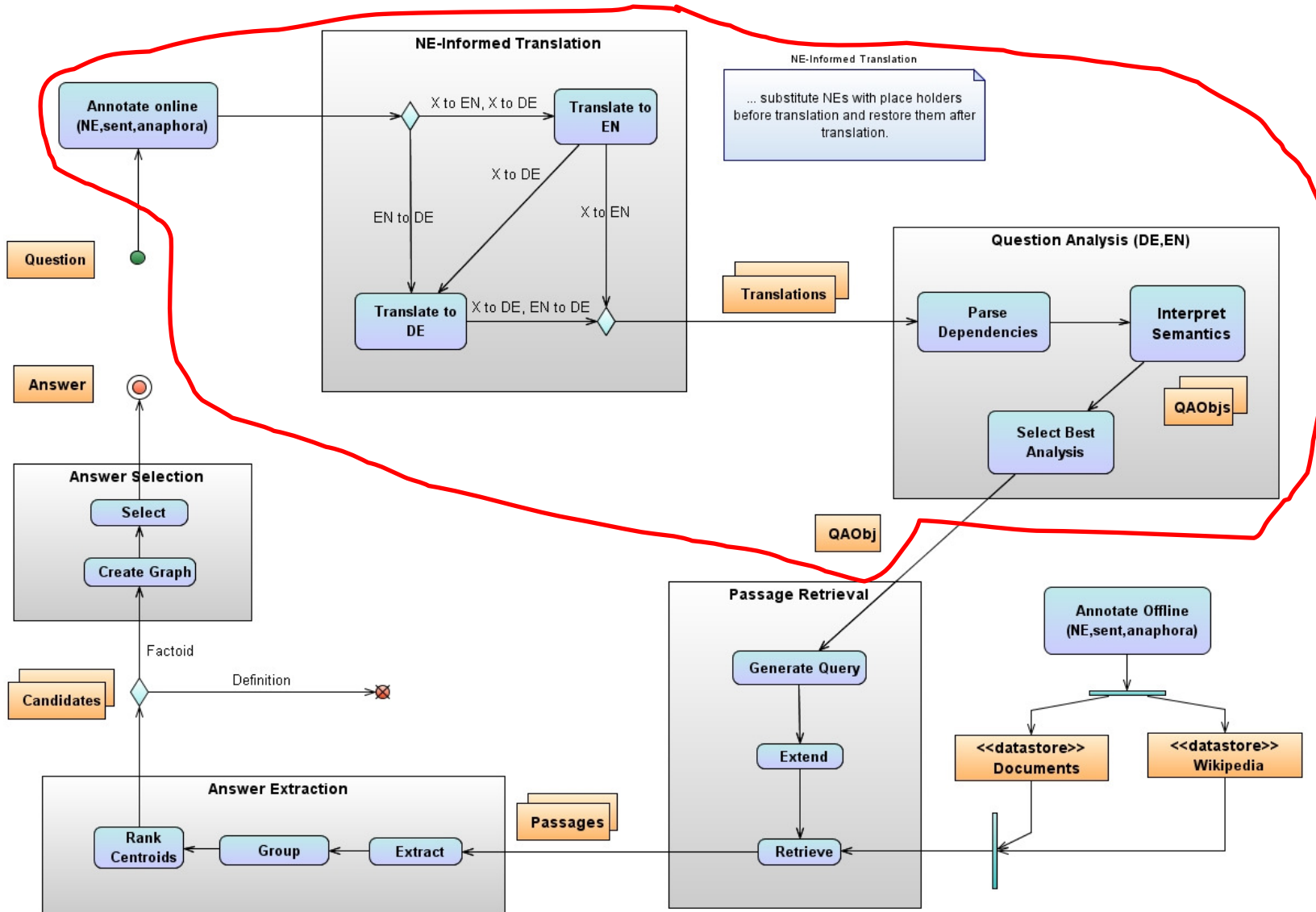
- Off-line: Provide question type oriented preprocessing for context selection
- On-line: Provide question specific preprocessing for answer processing



- ☆ Answer sources (covered by our technology)
  - Structured sources (DBMS)
  - Linguistically well-formed textual sources (news articles)
  - Well-structured web sources (Wikipedia)
  - Web snippets
  - Speech transcripts, cf. QAST
- ☆ Assumption:
  - QA for different answer sources share pool of same components
- ☆ Service oriented architecture (SOA) for QA
  - Strong component-oriented approach
  - Basis for open-source QA architecture (cf. EU project QALL-ME)









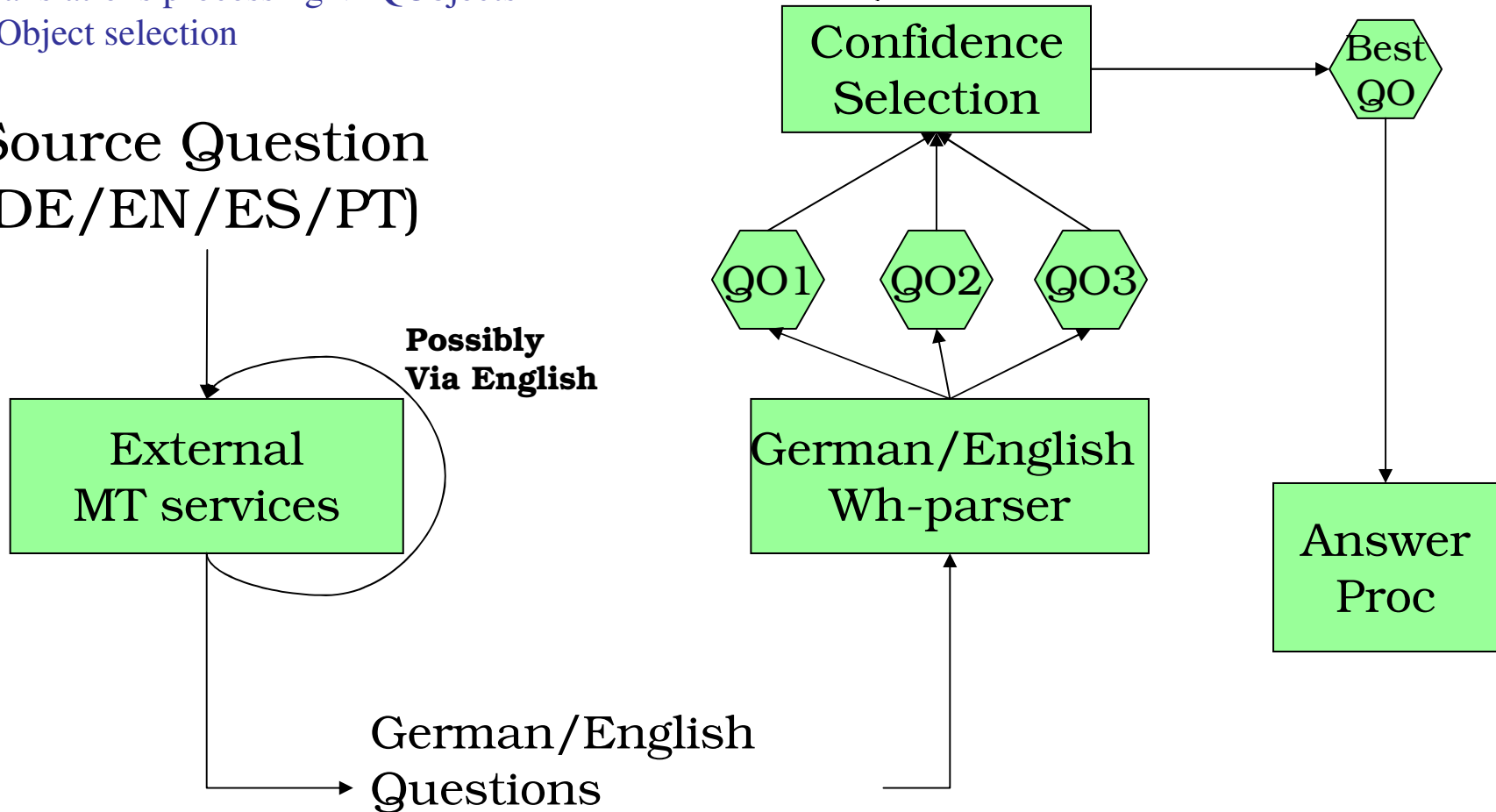
**Assumption: the better the query analysis of a translated question is done the better was the translation being made**

## Before Method

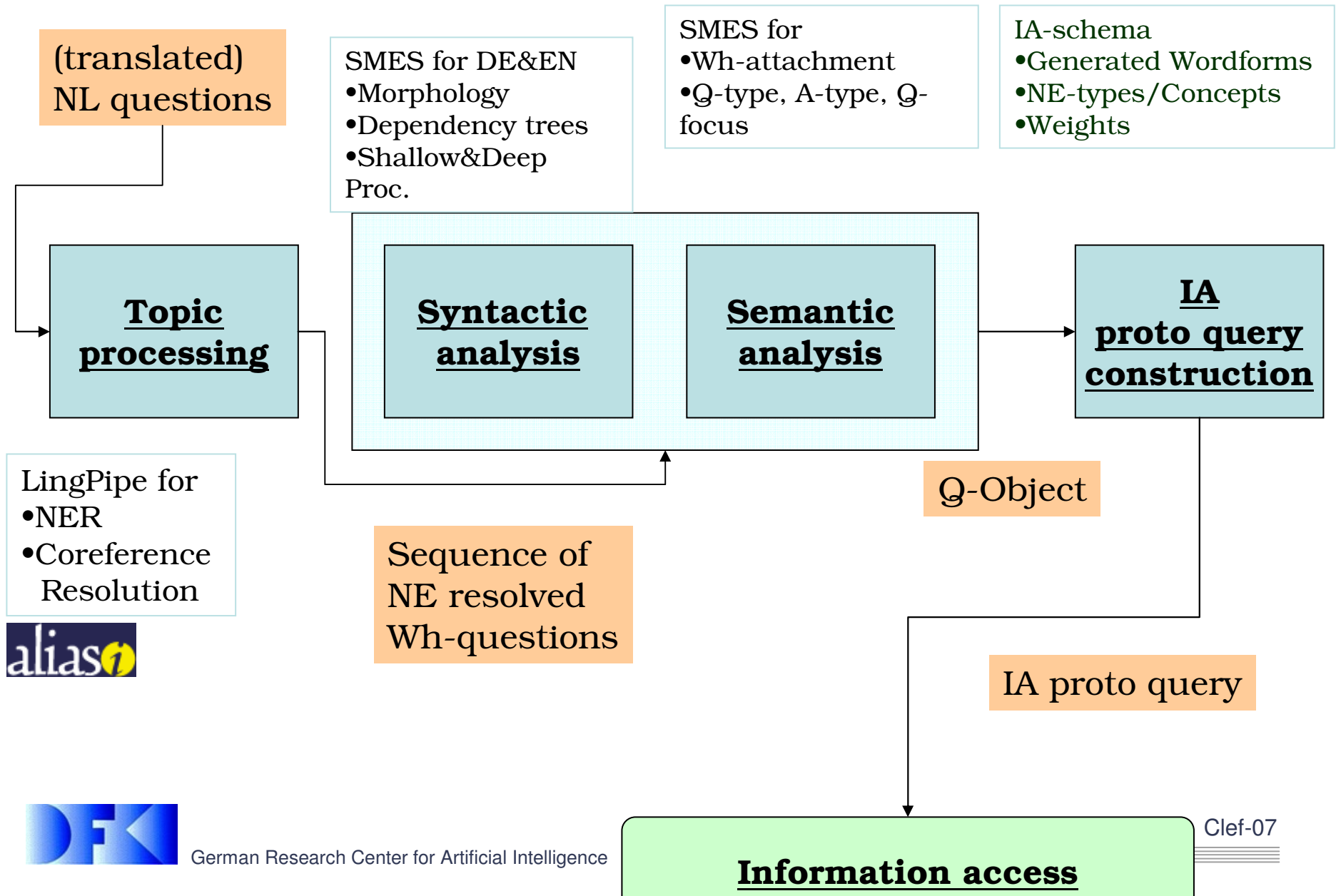
- Question translation
- Translations processing -> QObjects
- QObject selection

Completeness wrt.  
 -Parse tree  
 -major semantic Wh-types

Source Question  
 (DE/EN/ES/PT)









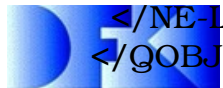
Which Jewish painter lived from

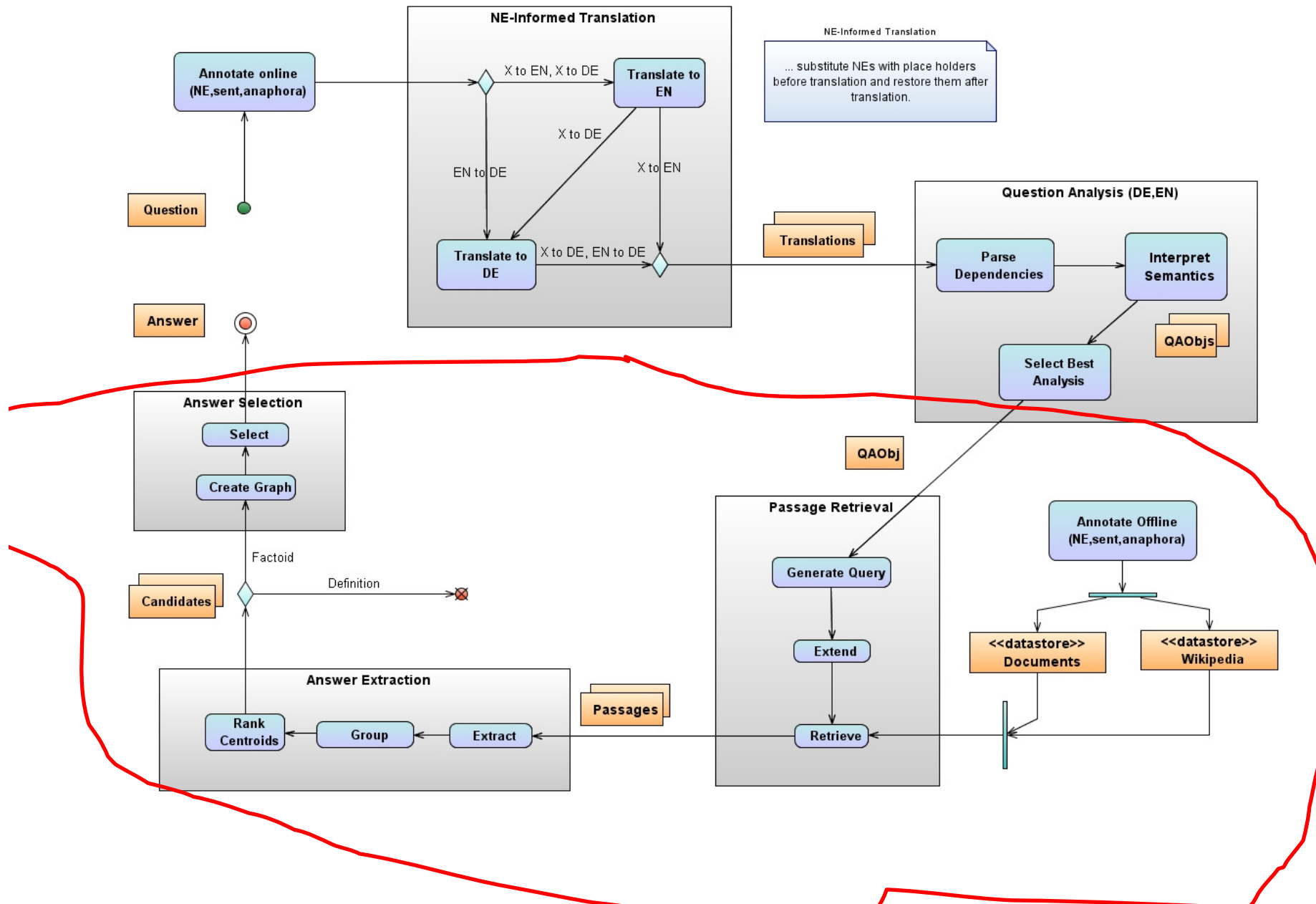
**Exploiting  
Natural Language  
Generation**

```
<QOBJ msg="quest" id="qId0" lang="DE" score
  <NL-STRING id="qId0">
    <SOURCE id="qId0" lang="DE">Welche juedische
Maler lebten von 1904-1944?</SOURCE>
  <TARGETS/>
</NL-STRING>
<QA-control>
  <Q-FOCUS>Maler</Q-FOCUS>
  <Q-SCOPE>leb</Q-SCOPE>
  <Q-TYPE restriction="TEMP">C-COMPLETION</Q-
TYPE>
  <A-TYPE type="list:SOME">NUMBER</A-TYPE>
</QA-control>
<KEYWORDS>
  <KEYWORD id="kw0" type="UNIQUE">
    <TK pos="V" stem="leb">lebten</TK>
  </KEYWORD>
  <KEYWORD id="kw1" type="UNIQUE">
    <TK pos="A" stem="juedisch">juedischen</TK>
    ...
</KEYWORD>
</KEYWORDS>
<EXPANDED-KEYWORDS/>
<NE-LIST>
  <NE id="ne0" type="DATE">1944</NE>
  <NE id="ne1" type="DATE">1904</NE>
</NE-LIST>
</QOBJ>
```

**IA query created for Lucene**

```
+neTypes:NUMBER
AND
("lebten" OR "lebte" OR "gelebt"
OR "leben" OR "lebt")
AND +maler^4
AND jüdisch^1
AND 1944^1
AND 1904^1
```







Run ID	Right		W	X	U
	#	%	#	#	#
<i>dfki061dede<sub>M</sub></i>	60	30	121	14	5
<i>dfki061ende<sub>C</sub></i>	37	18.5	144	18	1
<i>dfki061deen<sub>C</sub></i>	14	7	178	6	2
<i>dfki062esen<sub>C</sub></i>	10	5	180	10	0
<i>dfki062ptde<sub>C</sub></i>	5	2.5	189	4	2

**Performance still ok  
although some lost**

**Coverage problems of  
English Wh-parser**

**BUG in NE-Informed  
Translation (used DE-  
based recognizer)**

**Problems with MT  
online services  
(PT-EN-DE)**



- ☆ Online MT services are still insufficient
  - Develop own MT solutions, cf. EU project EuroMatrix
- ☆ Bad coverage of our English Wh-parser
  - First prototype for Clef 2007
- ☆ Answer extraction currently robust enough for different answer sources
  - Similar performance for newspaper and Wikipedia
- ☆ Need more semantic analysis on answer side without lost of coverage and domain-independency
  - We are exploring cognitive semantics (cf. Talmy, 1987)
- ☆ Number of QA components also used in QAST pilot task and AVE



☆ QAST pilot task

- For given written factoid question
- Extract answer from manual or automatic speech transcripts

☆ Answer Validation Exercise

- Given a triple of form (question, answer, supporting text)
- Decide whether the answer to the question is correct and
- Is supported or not according to the given supporting text

**Result (encouraging)**

Task	#Q	#A	MRR	ACC
T1	98	19	0.17	0.15
T2	98	9	0.09	0.09

T1 = Chill corpus manual  
 T2 = Chill corpus automatic

**Result (really encouraging)**

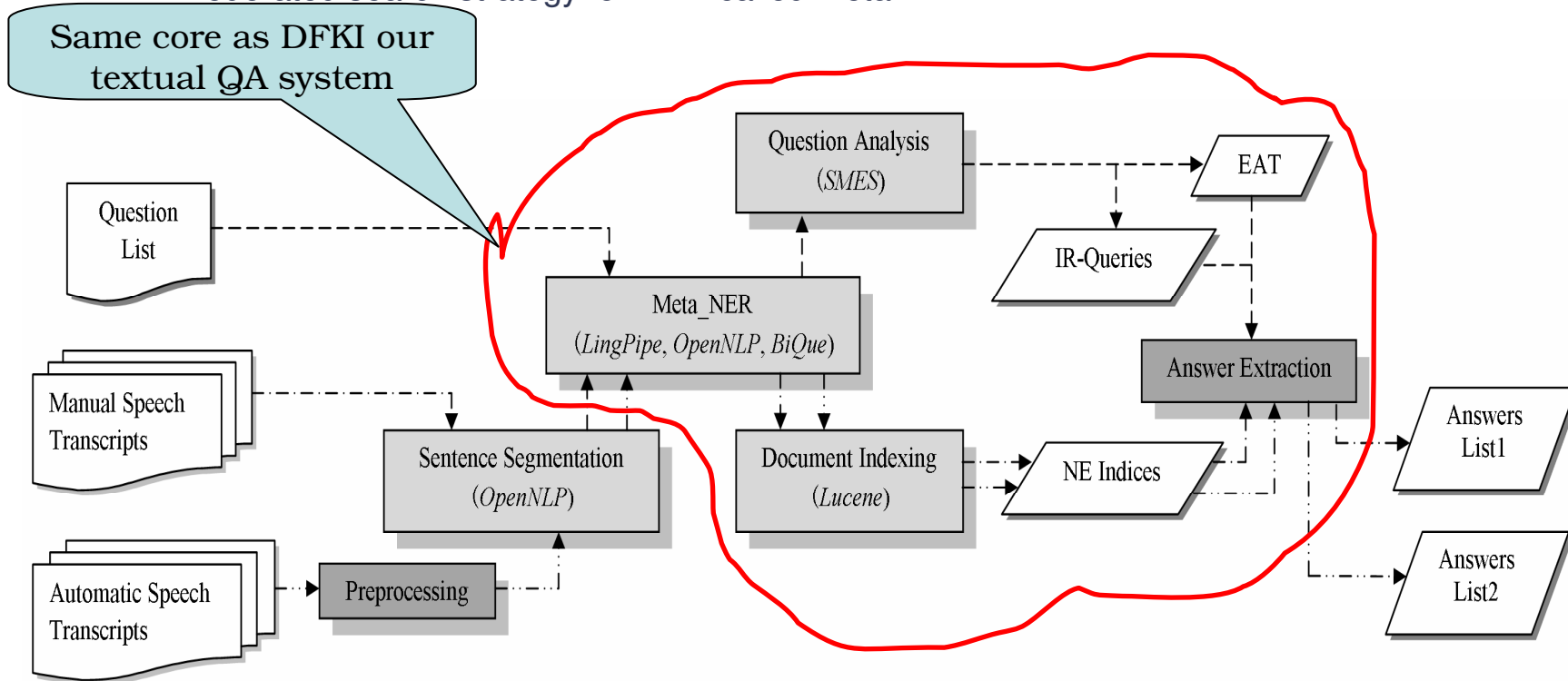
Runs	Recall	Precision	F-measure	QA Accuracy
dfki07-run1	0.62	0.37	0.46	0.16
dfki07-run2	0.71	0.44	<b>0.55</b>	0.21





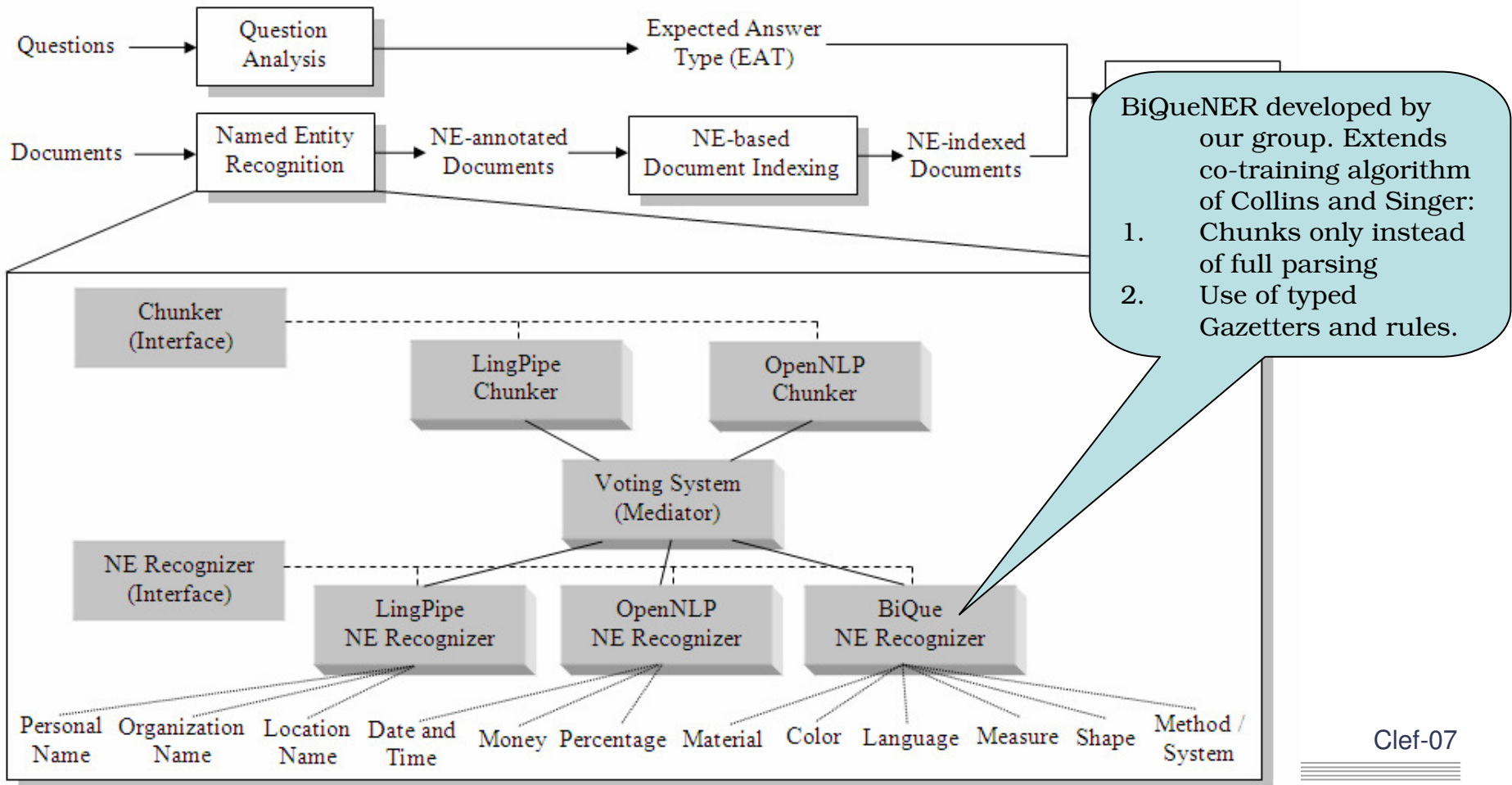
### ☆ Goals

- Get experience with this sort of answer sources
- Adapt our text-based open-domain QA system that we used for the Clef main tasks
- Since QAST required different set of expected answer types we developed a federated search strategy for NER called Meta-NER





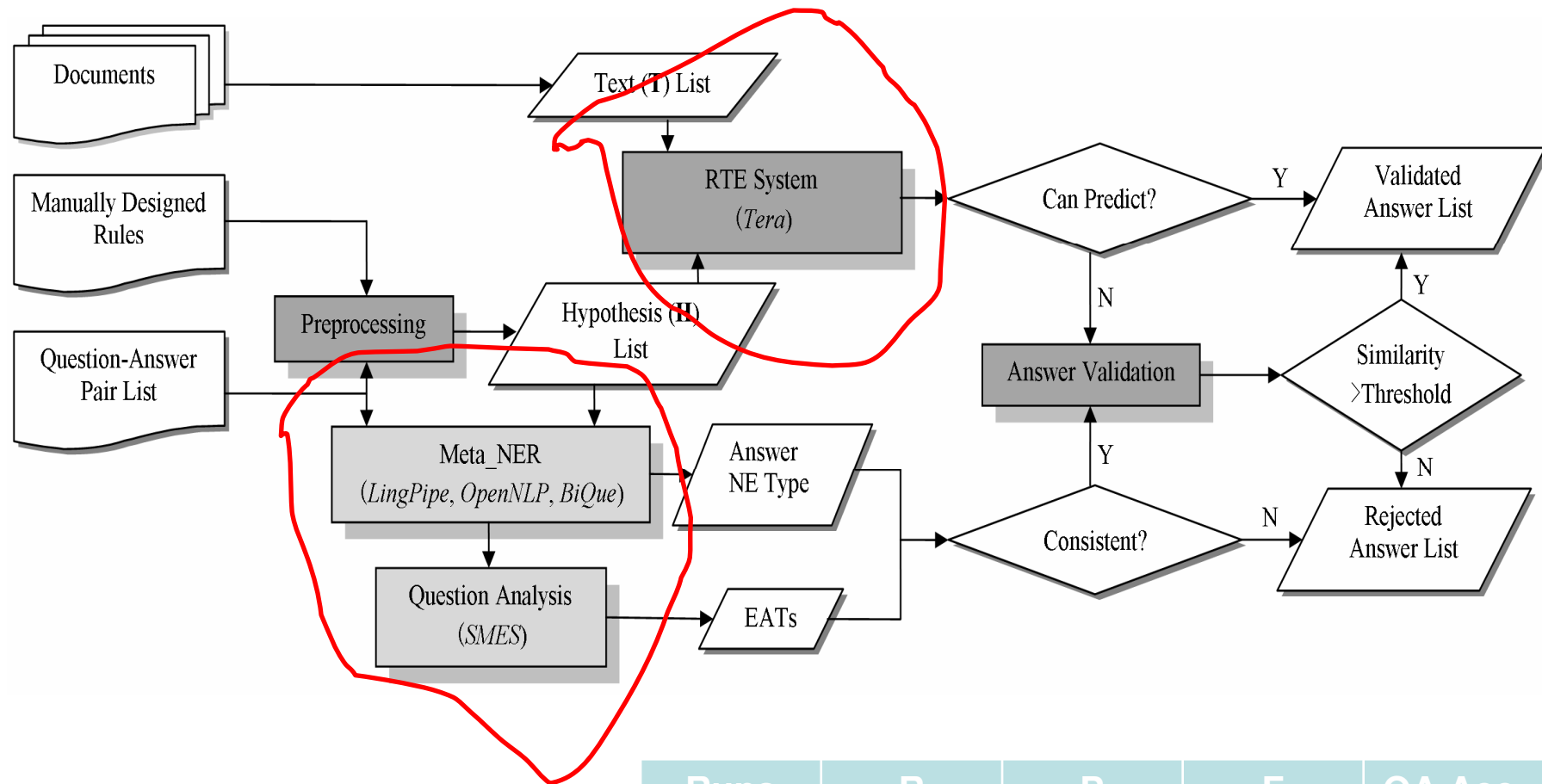
- ☆ Call several NER in parallel
- ☆ Merge results by a voting strategy







- ☆ AVE System is based on our RTE system (cf. Wang & Neumann, AAAI-2007, RTE-3 challenge)
- ☆ RTE method already demonstrated good results for QA task
  - RTE-3 (only QA): 81.5 %, Trec-2003 QA: 65.7 %
- ☆ RTE Method: Novel sentence level Kernel method
  - Subtree alignment on syntactic level
    - Check similarity between tree of H and relevant subtree in T
  - Subsequence kernel
    - Consider all possible subsequence of spine (path) of difference pairs
    - SVM for classification



Runs	R	P	F	QA Acc.
run1	0.62	0.37	0.46	0.16
run2	0.71	0.44	<b>0.55</b>	0.21





- ☆ Supporting text from web documents cause parsing problems
- ☆ Violation of some of our RTE system's assumptions
  - Required: H should be “verbally” smaller than T
  - Violated by: Q-A made patterns are too long
  - impact on recall
- ☆ If supporting text is very long (a complete document) then our RTE system is misled
  - Impact on precision



# **Thanks!**

---