Question Answering

Günter Neumann
Language Technology-Lab
DFKI, Saarbrücken

(with slides from Simon Sweeney)

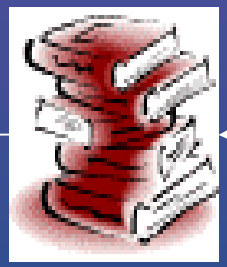
Motivation: From Search Engines to Answer Engines

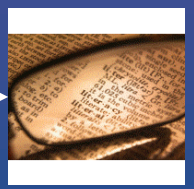
User Query:
KeyWrds, Wh-Clause, Q-Text

Search Engines

Experienc
ed-based
QA cycles

Answer Engines





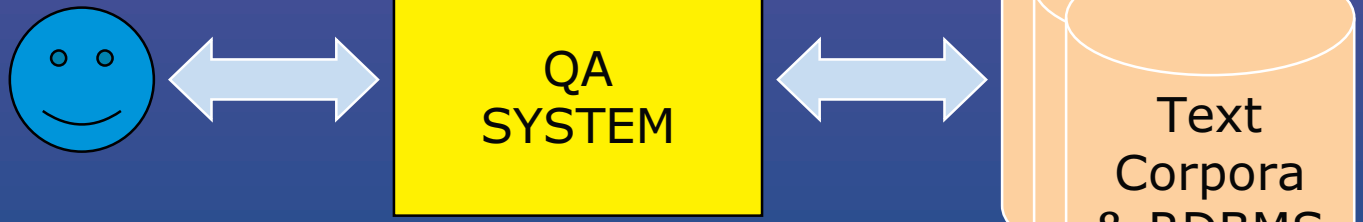
Shift more „interpretation effort“ to machines

User still carries the major efforts in understanding

Question Answering

- Input: a question in NL; a set of text and database resources
- Output: a set of possible answers drawn from the resources

"Where did Bill Gates go to college?"
"What is the rainiest place on Earth?"



"Harvard"

"Mount Waialeale"

"…Bill Gates, Harvard dropout and founder of Microsoft…" (Trec-Data)

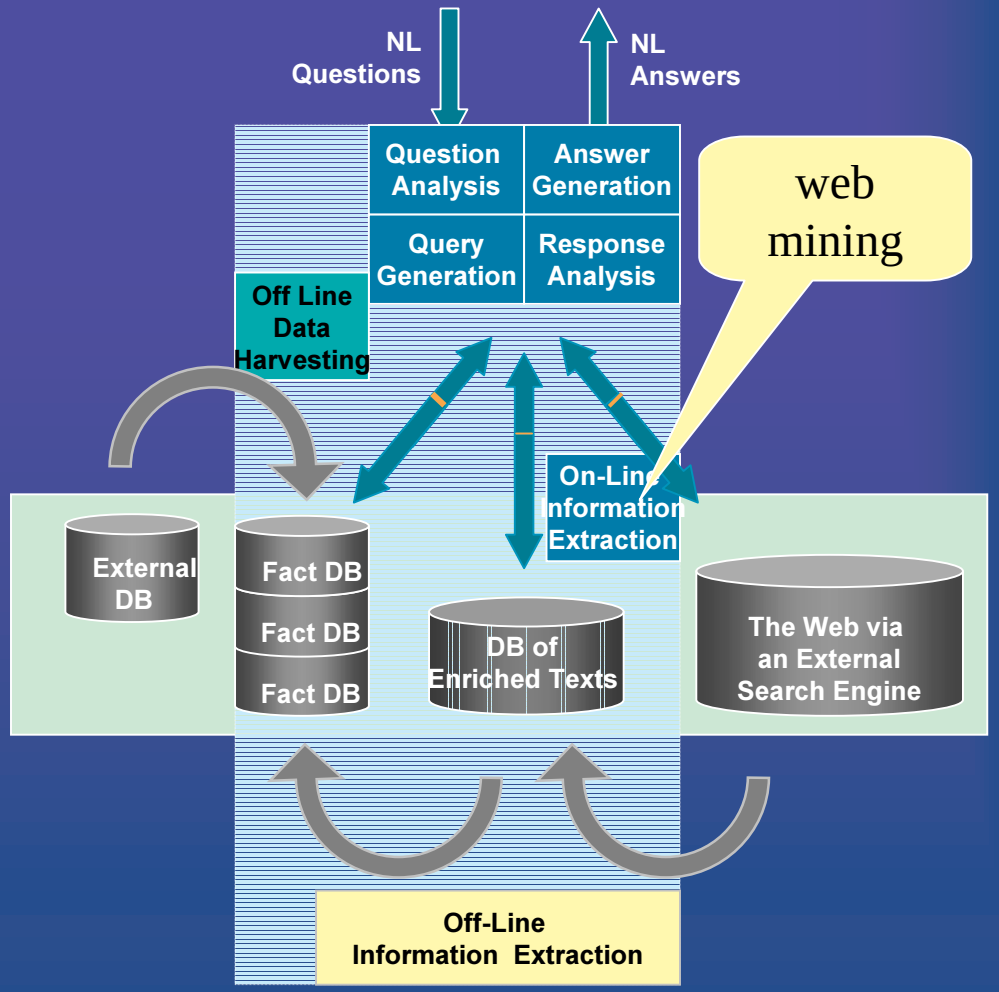
"… In misty Seattle, Wash., last year, 32 inches of rain fell.
Hong Kong gets about 80 inches a year, and even Pago Pago,
noted for its prodigious showers, gets only about 196 inches annually.
(The titleholder, according to the National Geographic Society,
is Mount Waialeale in Hawaii, where about 460 inches of rain falls each year.) …"
(Trec-Data; but see Google-retrieved Web page.)

Hybrid QA Architecture

Hypothesis

real-life QA systems will perform best if they can

- *combine* the virtues of domain-specialized QA with open-domain QA
- *utilize* general knowledge about frequent types and
- *access* semi-structured know-ledge bases



Advertisement:
DFKI project Quetal
2003-2005

Challenges for QA

- QA systems should be able to:
  - Timeliness: answer question in real-time, instantly incorporate new data sources.
  - Accuracy: detect no answers if none available.
  - Usability: mine answers regardless of the data source format, deliver answers in any format.
  - Completeness: provide complete coherent answers, allow data fusion, incorporate capabilities of reasoning.
  - Relevance: provide relevant answers in context, interactive to support user dialogs.
  - Credibility: provide criteria about the quality of an answer

Challenges for QA

- Open-domain questions & answers
- Information overload
  - How to find a needle in a haystack?
- Different styles of writing (newspaper, web, Wikipedia, …)
- Scalability & Adaptibility

Information Overload

"The greatest problem of today is how to teach people to ignore the irrelevant, how to refuse to know things, before they are suffocated. For too many facts are as bad as non at all". (W.H. Auden)

Problems in Information Access?

- Why is there an issue with regards to information access?
- Why do we need support in find answers to questions?

- IA increasingly difficult when we have consider issues such as:
  - the size of collection
  - the presence of duplicate information
  - the presence of misinformation (false information)

Traditional Information Access

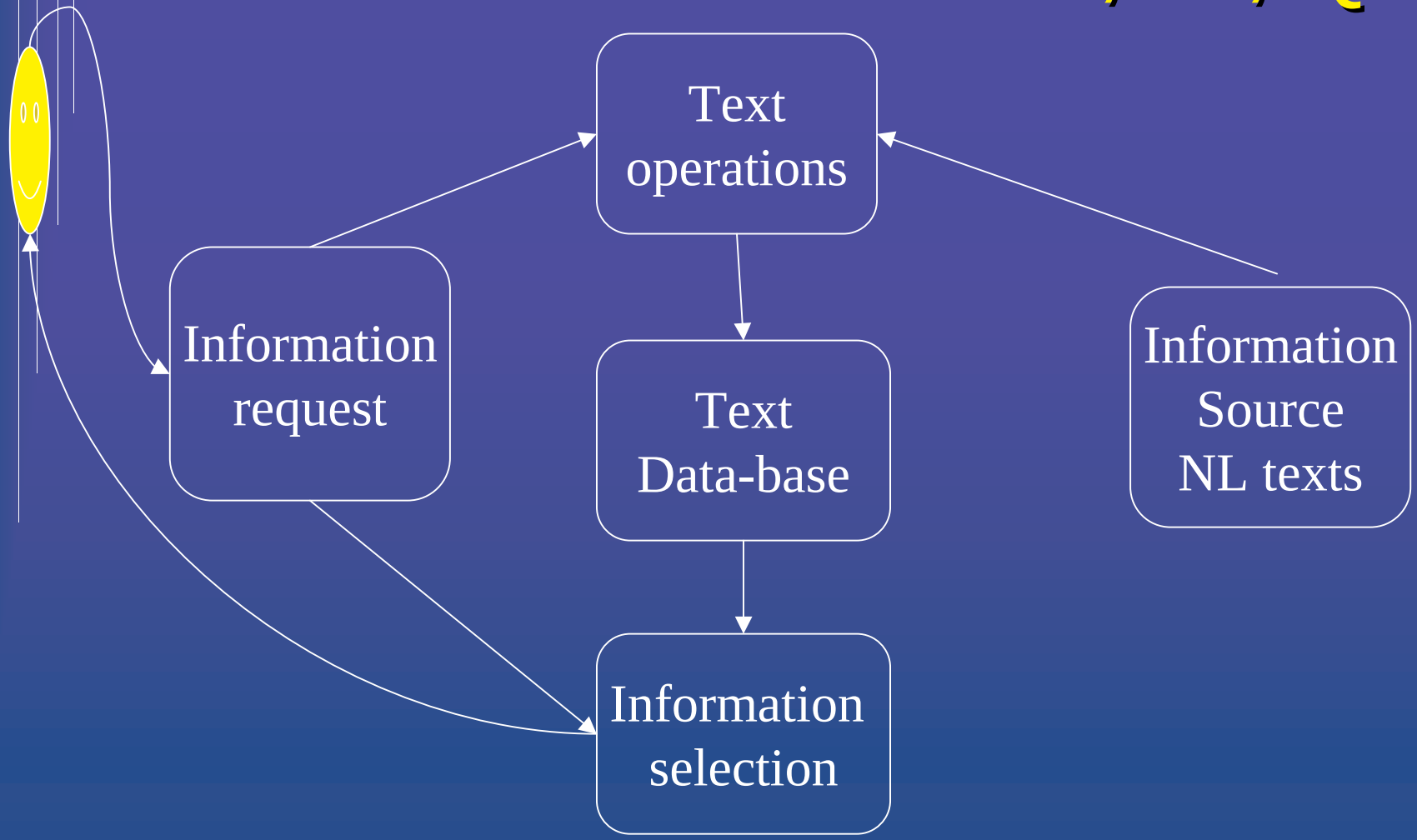
- Information Retrieval (IR)
- Information Extraction (IE)

- Question Answering (QA) (new)

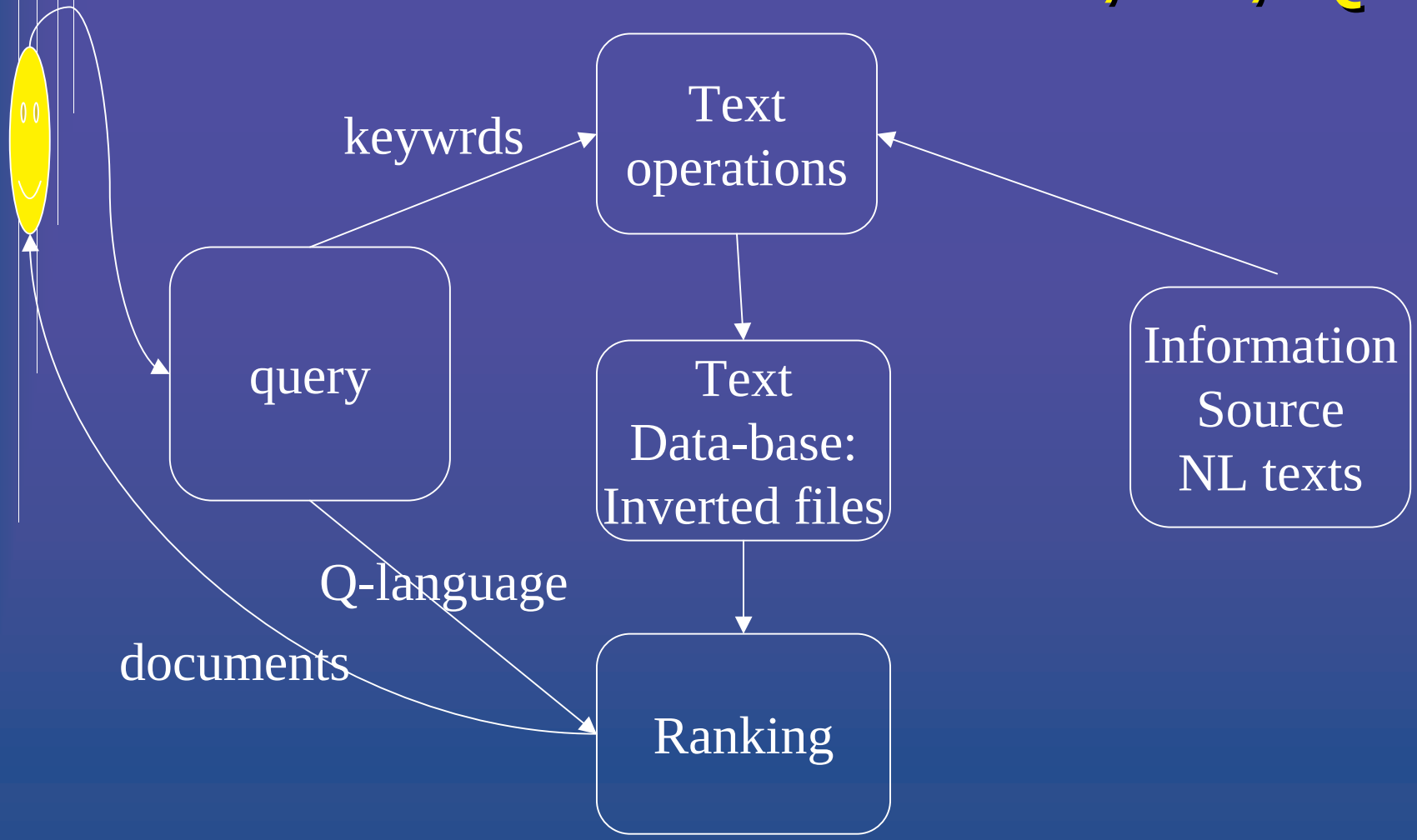
What is Question Answering ?

- Natural language questions, not queries
- Answers, not documents (containing possibly the answer)
- A resource to address ‘information overload’?
- Current research as focused on fact-based questions:
  - “How tall is Mount Everest?”,
  - “When did Columbus discover America?”,
  - ”Who was Grover Cleveland married to?”.
- One goal in QA is to eventually support information-seeking dialogs:
  - “Do you mean President Cleveland?”
  - “Yes”.
  - “Francis Folsom married Grover Cleveland in 1886.”
  - What was the public reaction to the wedding?

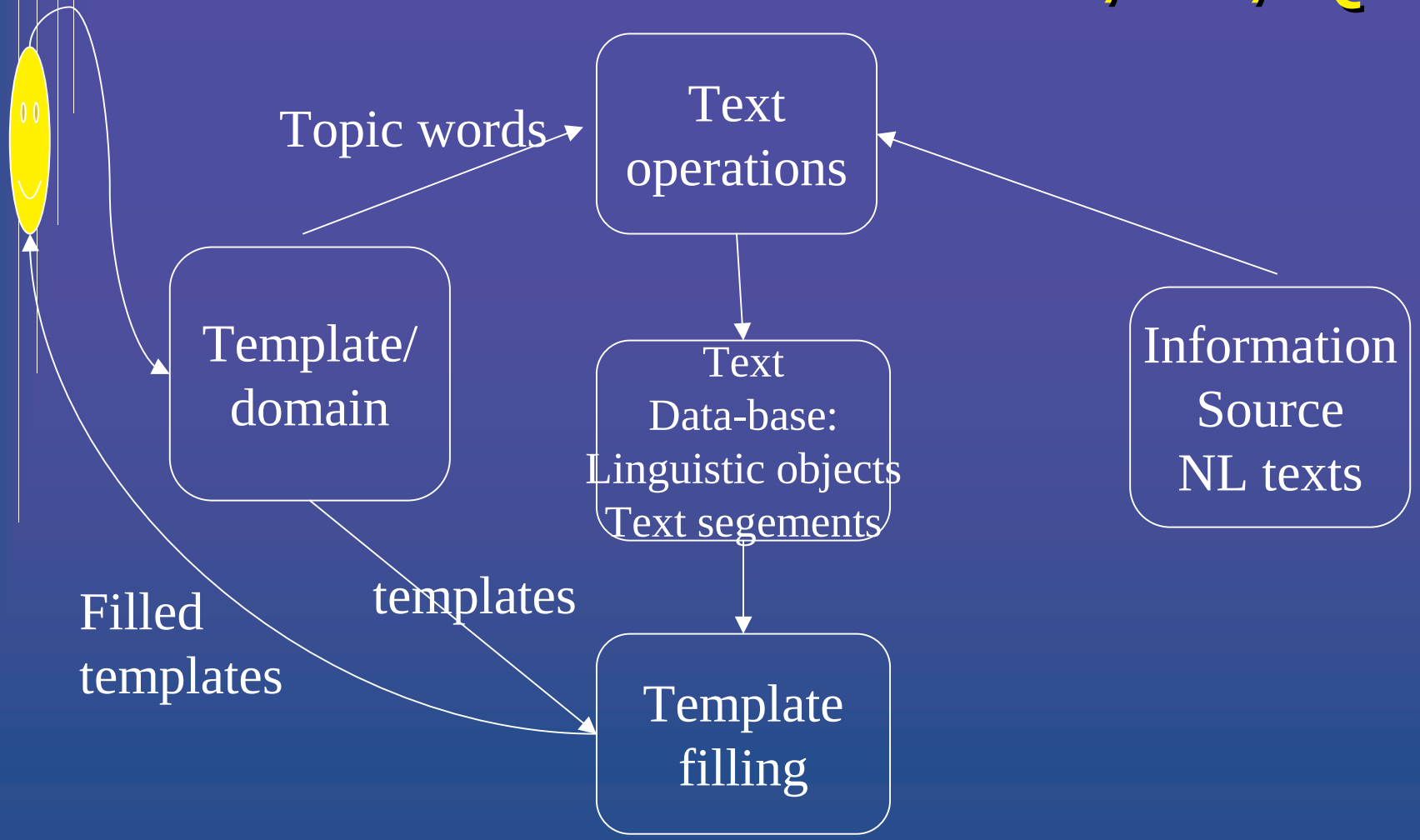
A Common View on IR, IE, QA



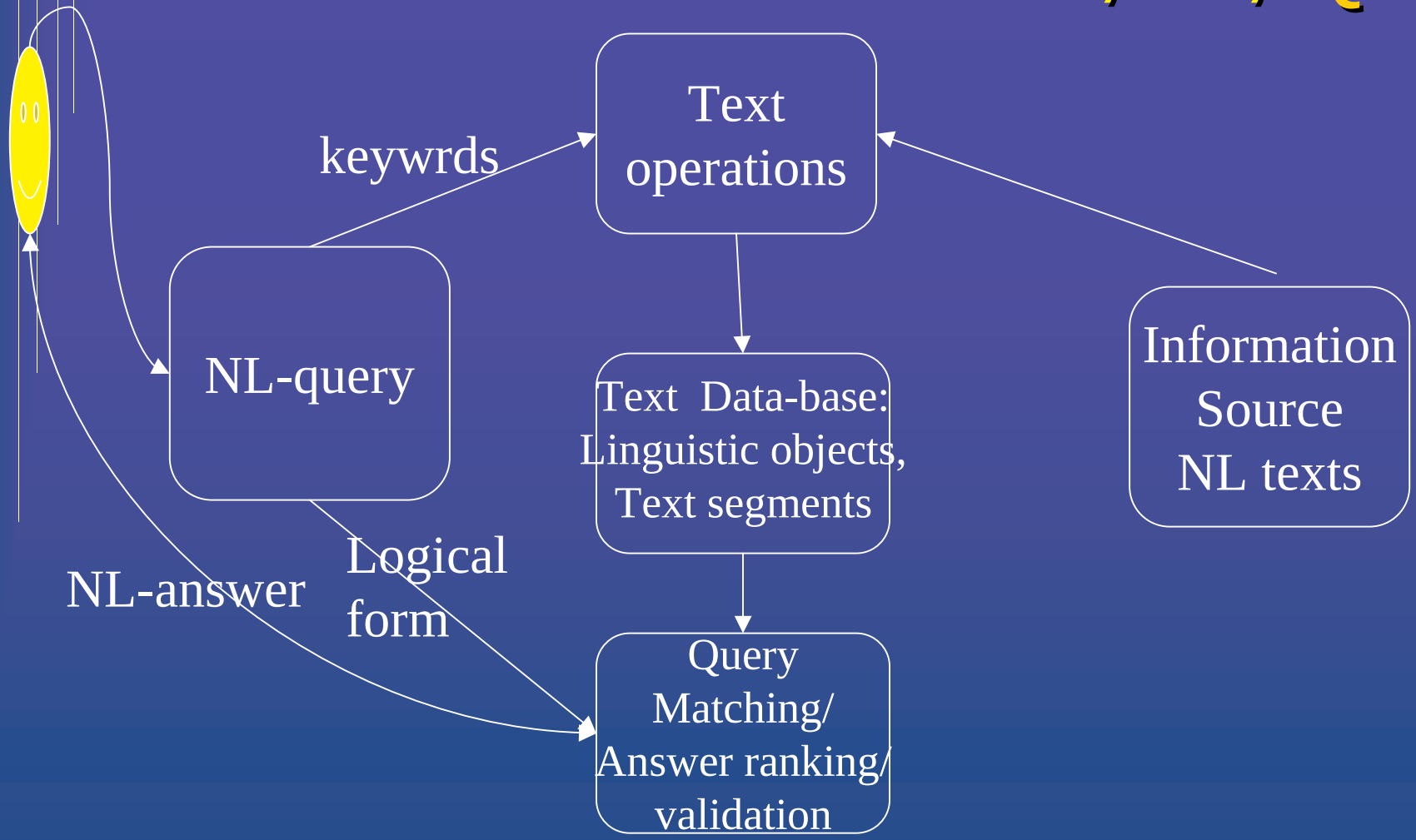
A Common View on IR, IE, QA



A Common View on IR, IE, QA



A Common View on IR, IE, QA



Brief history of QA Systems

- QA has had contributions from a number of fields, such as,
  - Database: NL front ends to databases
    BASEBALL (1961), LUNAR (1973)
  - AI: dialog interactive advisory systems
    SHRLDU (1972), JUPITER (2000)
  - NLP: story comprehension
    BORIS (1972)
  - NLP: retrieved answers from an encyclopedia
    MURAX (1993)
- TREC's QA track (began in 1999)

TREC QA track

- What is TREC?
  - Text REtrieval Conference is a series of workshops aim at developing research on technologies for IR.
  - started: 1992, Sponsored by: NIST, DARPA
  - TREC-10 (2001), no. of tracks: 6, no. participants: 87
- What is TREC QA track?
  - focuses on the evaluation of systems, in a competition-based manner, that answer questions in unrestricted domains.
  - started: TREC-8 (1999), no. participants: 20

TREC-8,-9,-10 QA tracks

- QA Track
  - NIST provides: a document collection, and a set of test questions.
  - Participants build a QA system and return

	TREC-8	TREC-9	TREC-10
no. of documents	528,000	979,000	979,000
megabytes of document text	1904	3033	3033
no. of questions	200	693	500*
document sources	newspaper articles	expanded to incl. AP newswire, WSJ	expanded to incl. AP newswire, WSJ
question sources	(FAQ Finder log) assessors	Encarta log, Excite log	MSNSearch, AskJeeves log
question type	fact-based short answer	more difficult questions	3 tasks: main task, list task, context task

* 135 definitional questions

TREC QA track results

- Lessons learnt from early track results:
  - QA works!
- Some results:

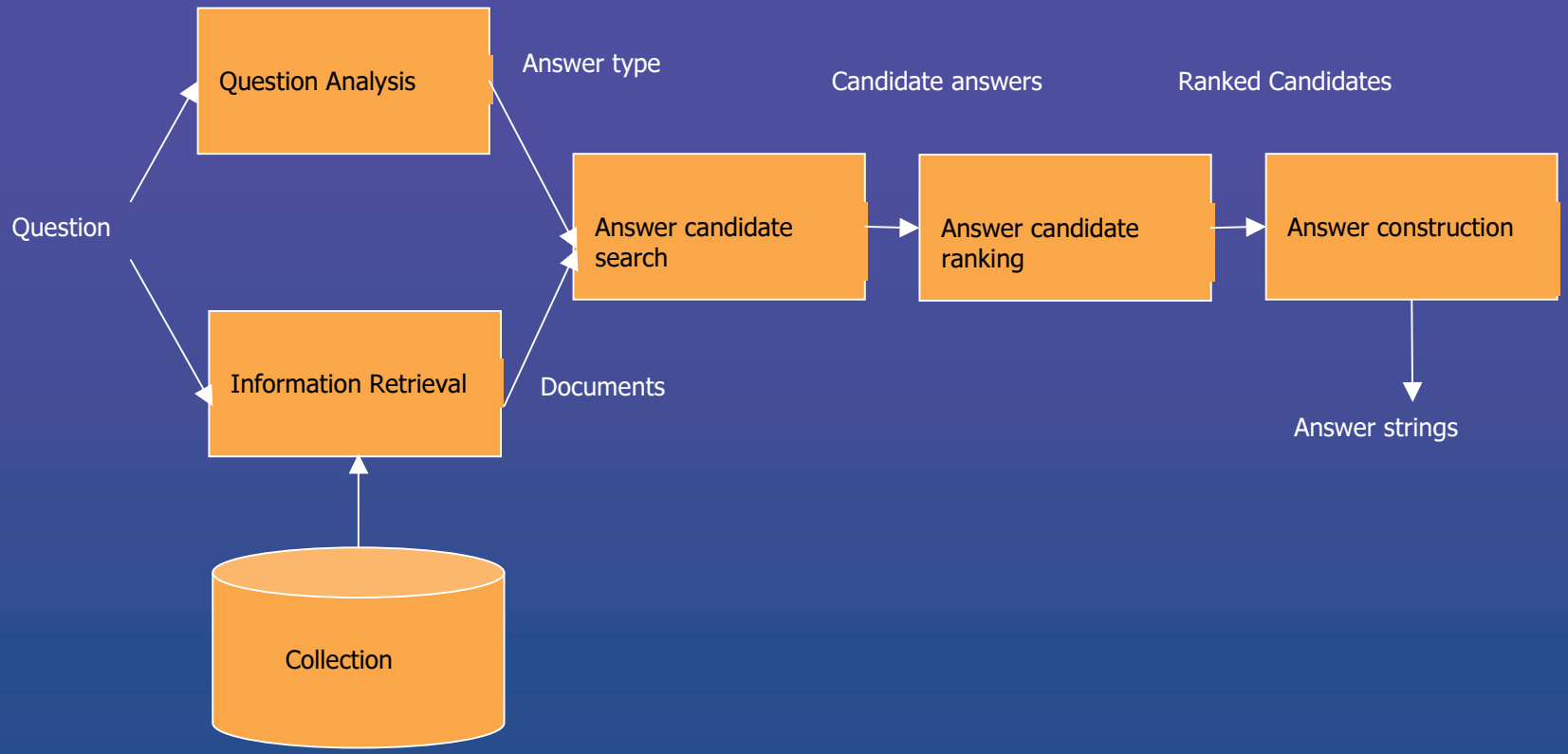
	Top performing QA System	MRR (%)
TREC-8	Cymfony	67*
TREC-9	LCC (SMU)	76*
TREC-10	InsightM-soft	69

*Short answer category, 50-byte

- TREC-10 QA, top 3 systems:
  - InsightSoft-M: TextRoller System (MRR 69%)
  - Language Computer Corporation: QAS System* (MRR 59%)
  - Oracle: QA System (MRR 49%)

*competed in all 3 tasks.

A General QA system Architecture



General Approach to QA

- Observations from TREC systems:
  - Stage 1: Question Analysis

    Find type of object that answers question: “when” – time, date

    “who” – person, organisation….
  - Stage 2: Document Retrieval

    Using (augmented) question, retrieve set of possible relevant document

    using IR.
  - Stage 3: Document processing

    Search documents for entities of the desired type using IE.

    Search entities in appropriate relations.
  - Stage 4: Rank answer candidates
  - Stage 5: Answer construction

Design Issues

- Foster bottom-up system development
  - Data-driven, robustness, scalability
  - From shallow & deep NLP

- Large-scale answer processing
  - Coarse-grained uniform representation of query/documents
  - Text zooming
    - From paragraphs to sentences to phrases
  - Ranking scheme for answer selection

- Common basis for
  - Online Web pages
  - Large textual sources

Open-Domain Question Answering

- Open domain
  - No restriction for the domain and type of question
  - No restriction on document source
- Combines
  - Information retrieval
  - Information extraction
  - Text mining
  - Computational Linguistics
- Cross-lingual ODQA
  - Express query in language X
  - Answer from documents in language Y

Open-Domain Question-Answering

"Mit wem ist David Beckham verheiratet?"

{person:David Beckham, married, person:?}

German Question

Question Analysis

English Question Object

IR-Query Construction

IR-Google

IR-Lucene/XML

Annotated Corpus

Query Translation

- Online MT-systems
- WSD
- Expansion

Documents

Question Object:

- Focus, Scope
- AnswerType

Passage selection

Passages

"David Beckham, the soccer star engaged to marry Posh Spice, is being blamed for England 's World Cup defeat."

Answer

Answer Selection

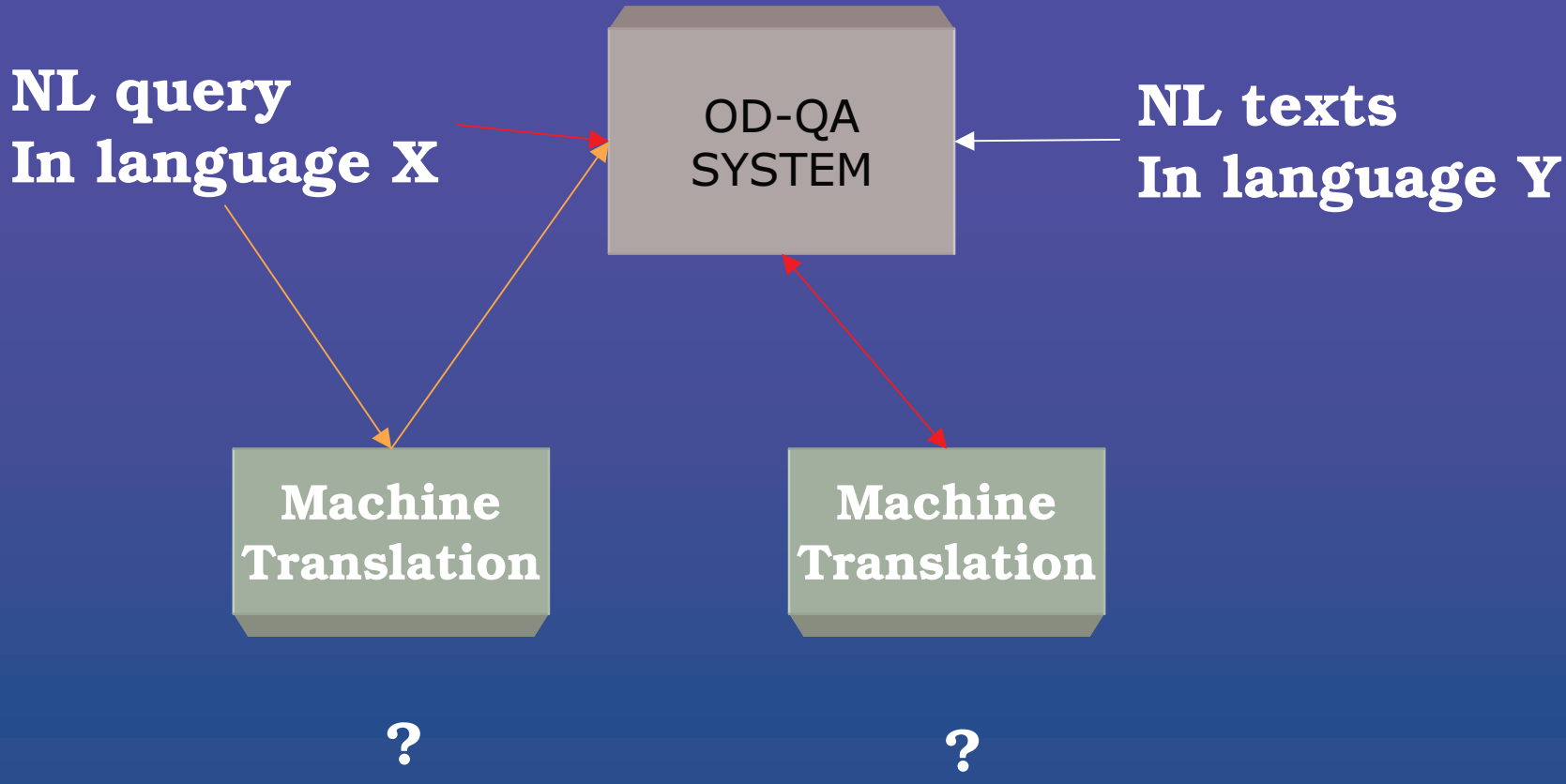
Candidates

Answer Extraction

Posh Spice

{person:David Beckham, person:Posh Spice}

Cross-Lingual ODQA



Approaches in CL QA

Two main different approaches used in Cross-Language QA systems:

1

translation of the question into the target language (i.e. in the language of the document collection)

question processing

answer extraction

2

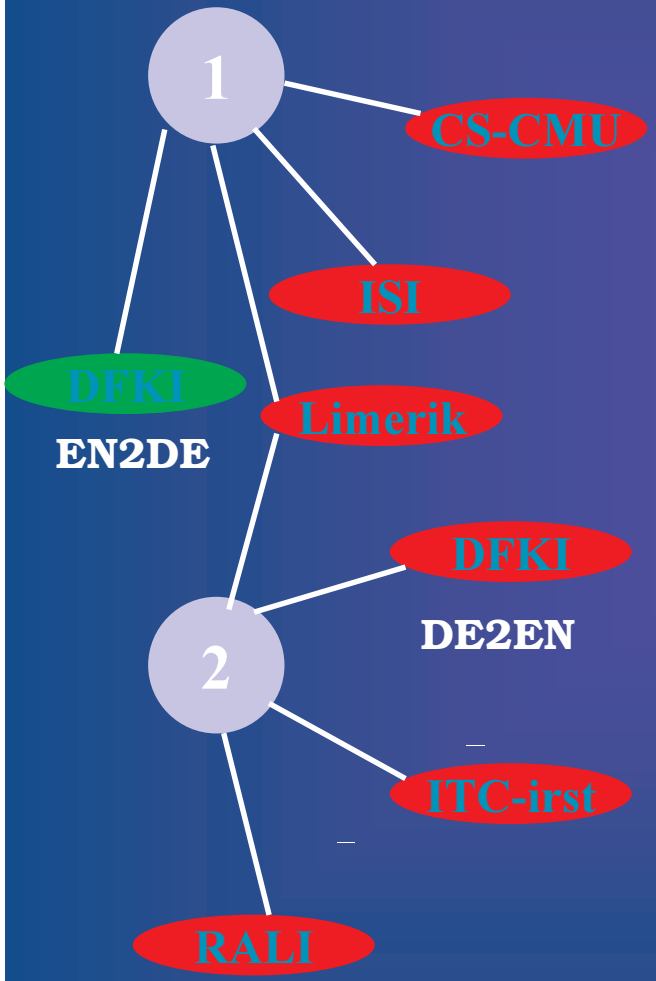
question processing in the source language to retrieve information (such as keywords, question focus, expected answer type, etc.)

translation and expansion of the retrieved data

answer extraction

Approaches in CL QA

Two main different approaches used in Cross-Language QA systems:



translation of the question into the target language (i.e. in the language of the document collection)

question processing

answer extraction

question processing in the source language to retrieve information (such as keywords, question focus, expected answer type, etc.)

translation and expansion of the retrieved data

answer extraction

Query Translation & Expansion

- At DFKI we tried the following:
  - Only use EuroWordNet, a multilingual thesarus
  - Defines a word-based translation via synset offsets
- Experience
  - EuroWordNet too sparse on German side
  - Neverless introduced too much ambiguity
  - Translation of Named Entities is crucial
- So far, not very much of help

The EuroWord Database

- Multilingual thesaurus containing 8 languages (as to 2001)

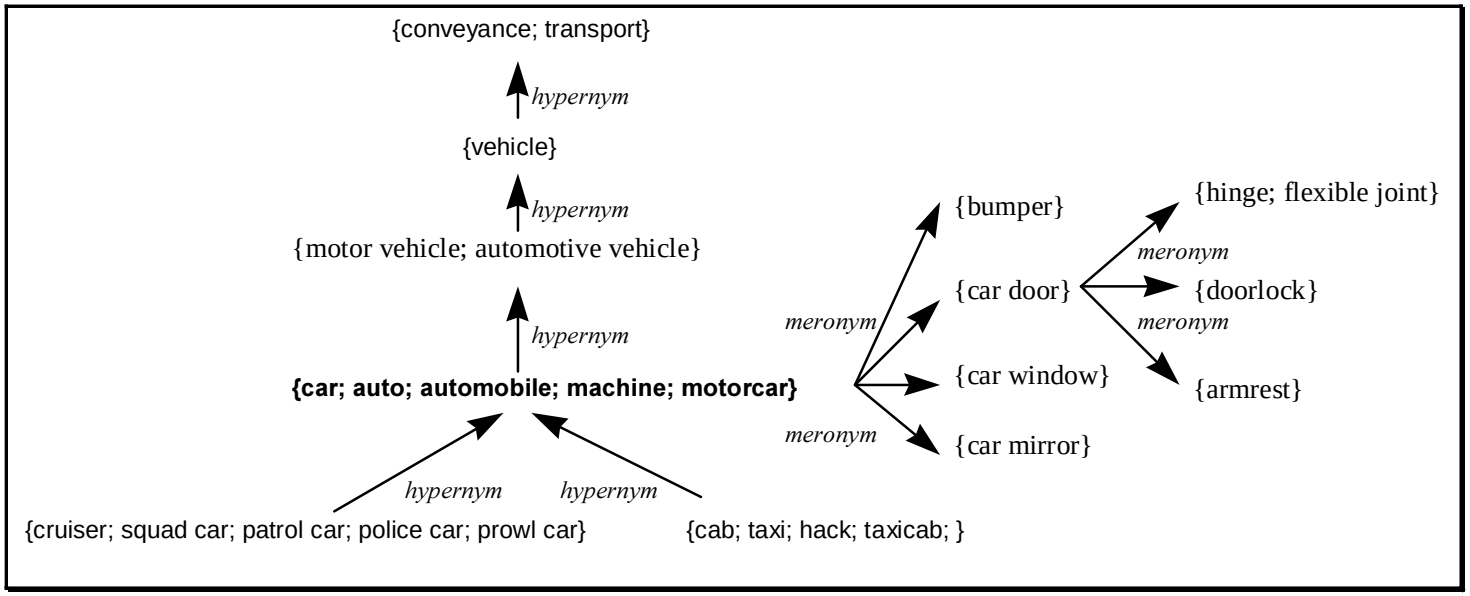
Language	No. of words
English	120,000
Dutch	55,000
Spanish	32,000
Italian	32,000
German	17,000
French	19,000
Estonian	11,000
Czech	12,000

- Under development:
  Swedish, Norwegian, Danish, Greek, Portuguese, Basque, Catalan,
  Romanian, Lithuanian, Russian, Bulgarian, Slovenian.

Basis of EuroWordNet

- Based on WordNet (http://www.cogsci.princeton.edu/~wn)
- On-line English lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory.
- English nouns, verbs, adjectives and adverbs are organized into synonym sets (**synsets**), each representing one underlying lexical concept.
- Different semantic relations link the synonym sets.

WordNet Synsets and Relations



WordNet Relations

nouns

antonymy		foe - friend
synonymy		jester - fool
hyponymy		vertebrate - bird
hypernymy		parrot - bird
holonymy	member/part/substance	door -wall
meronymy	member/part/substance	bread - flour
attribution		size - large
pertainymy		pole - polar

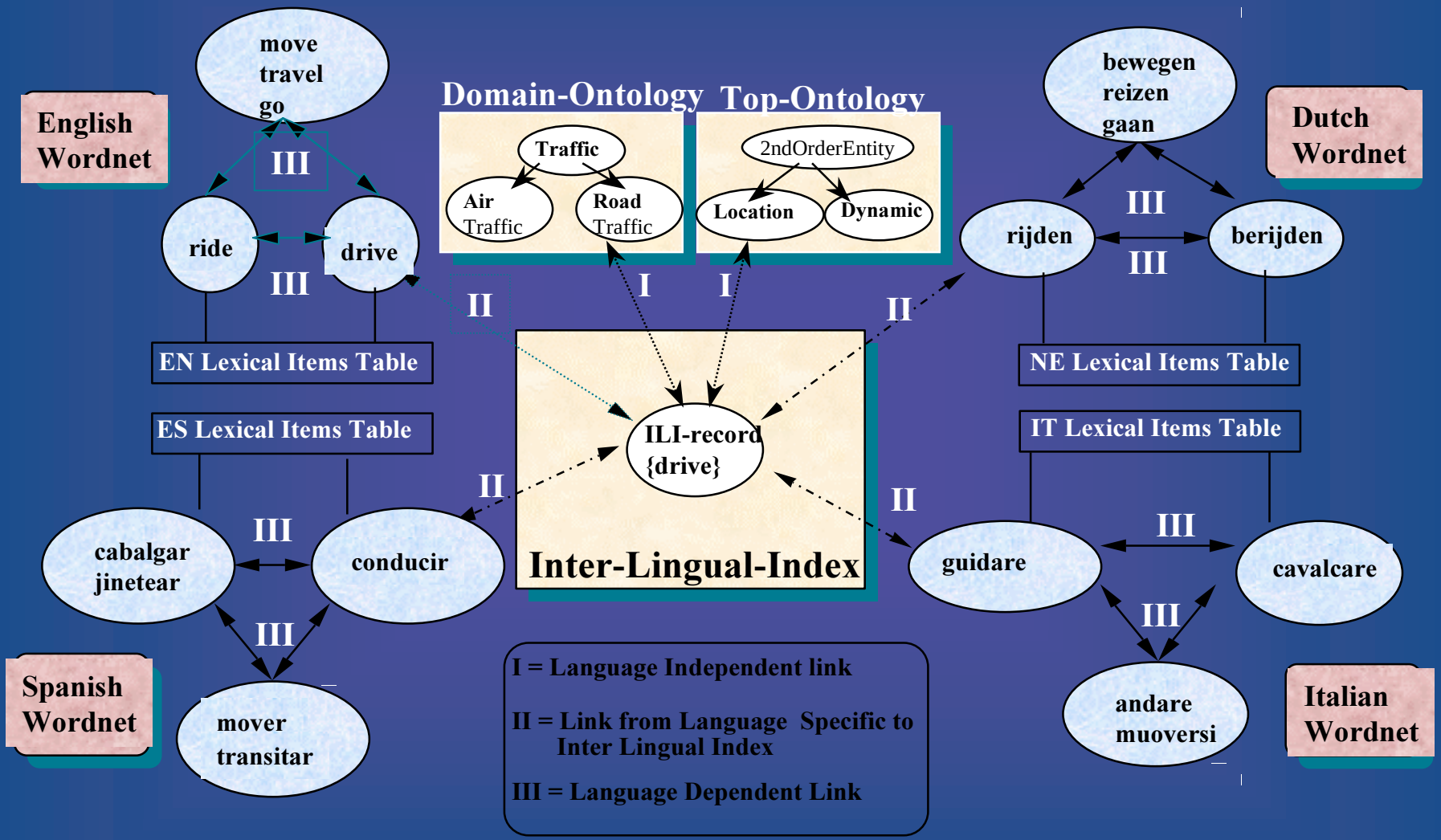
verbs

synonymy		proceed - go
antonymy		come - go
hypernymy		go - do something
hyponymy		go - travel
entailment		buy - pay
causation		break - fall apart

Structure of EuroWordNet

- Language-specific independent wordnets
- Wordnets represent unique concept lexicalization patterns in 8 languages, based on sense-inventories of mono- and bilingual dictionaries
- Concepts are synsets, i.e. sets of synonyms
- fixed set of internal relations between concepts
- These are linked to the Inter-Lingual Index (ILI) which serves as the interlingua
- ILI is based on the synsets from WordNet1.5
- Most synsets (85%) are directly linked to the closest concepts in the Inter-Lingual-Index by means of equivalence relations

Architecture of the EuroWordNet Data Base

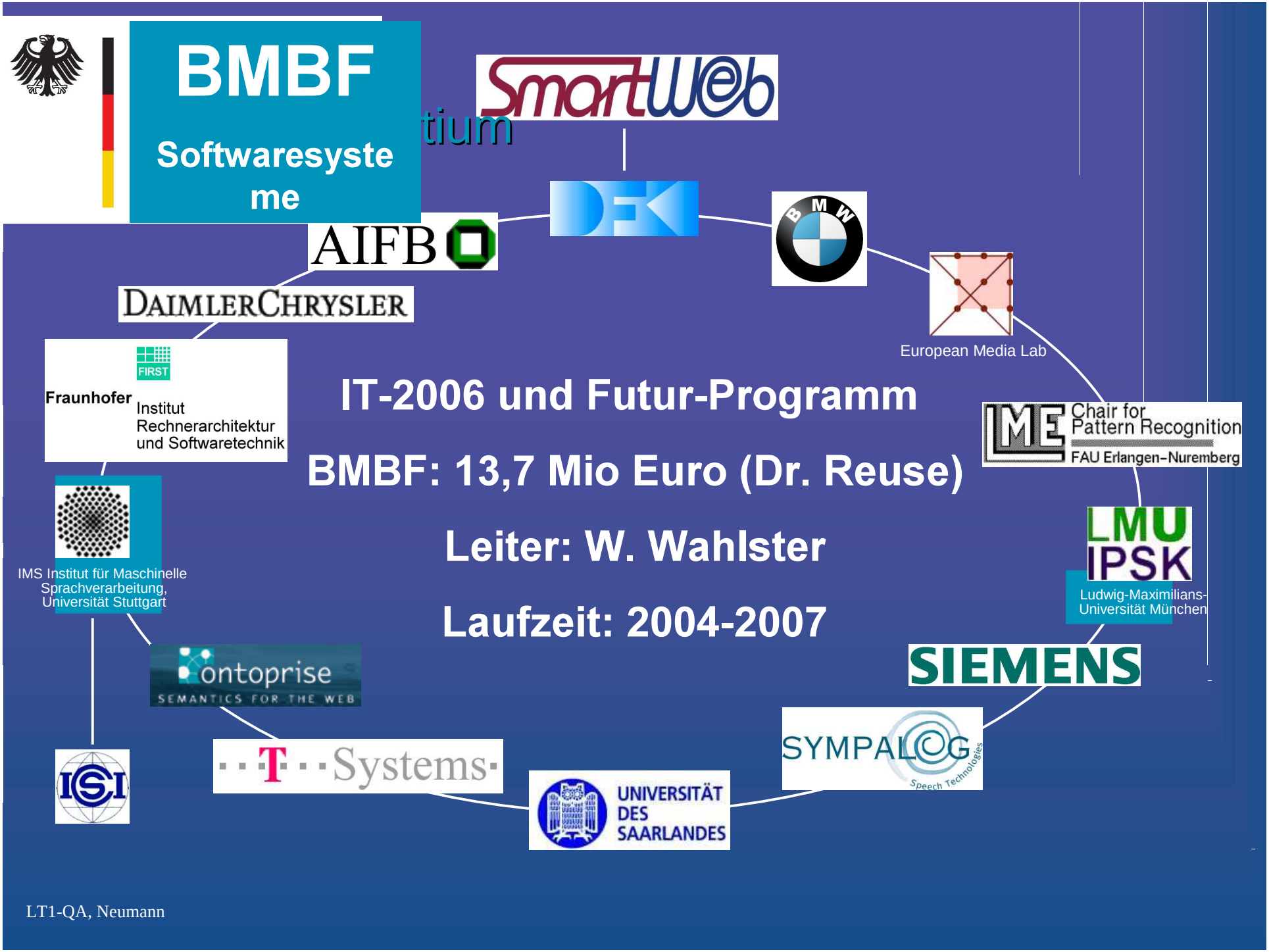


QA@DFKI LT-lab

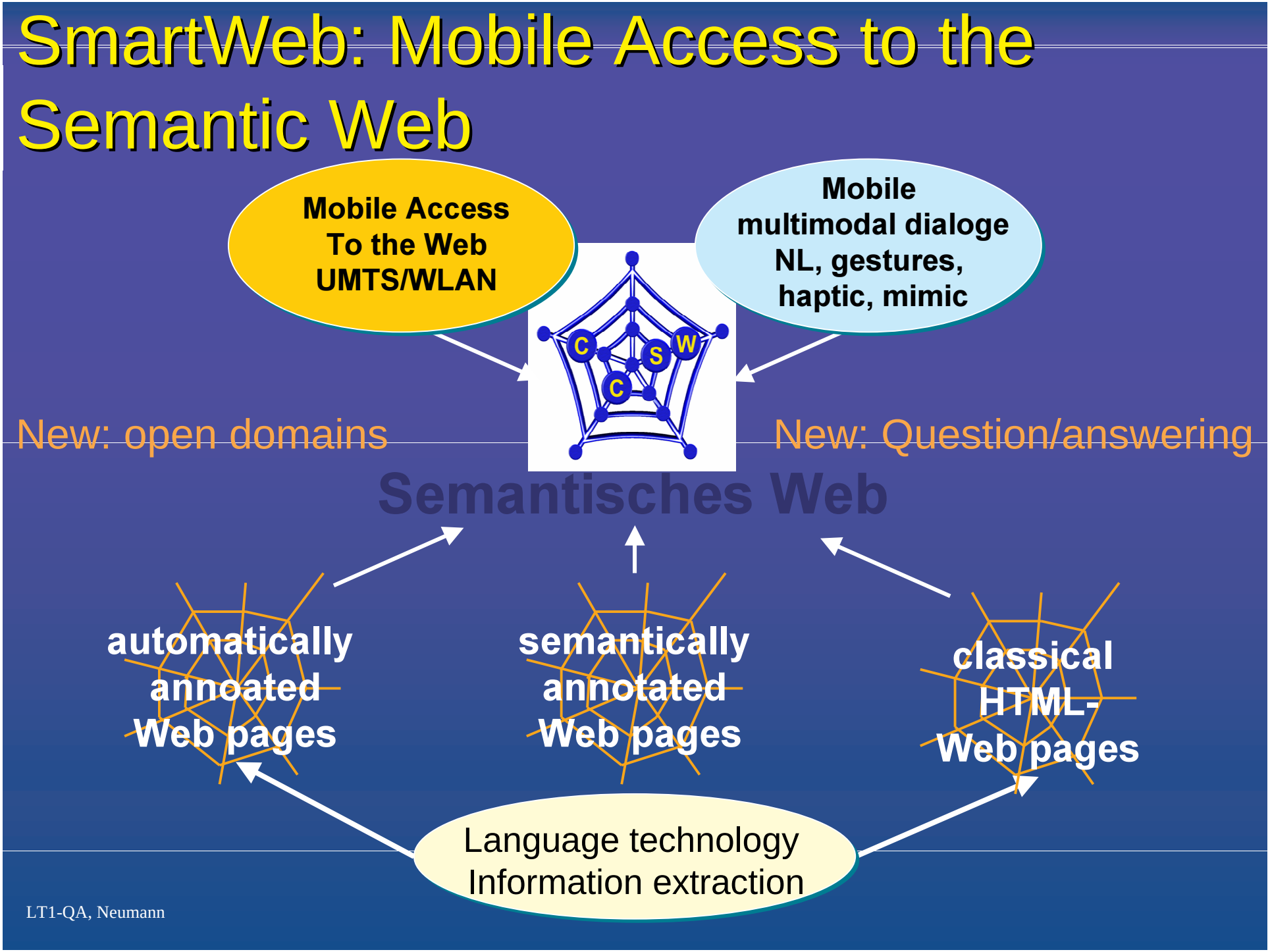
- Cross-lingual open-domain QA
- QA for the semantic web
- QA for the personal digital memory

DFKI projects

- Quetal
  - Hybrid QA
  - Cross-lingual open-domain QA
- SmartWeb
  - Mobile access to the Semantic Web
  - ODQA for search in syntactic Web pages
  - BMBF Verbund project (partners from research and industries, e.g., BMW, Siemens, T-Systems)
- HyLab
  - ODQA for Personal Digital Memory
  - BMBF funded DFKI project starting next year


BMBF
Softwaresyste
me
SmartWeb
tium
AIFB
DEK
BMW
DAIMLERCHRYSLER
European Media Lab
FIRST
Fraunhofer
Institut
Rechnerarchitektur
und Softwaretechnik
IT-2006 und Futur-Programm
BMBF: 13,7 Mio Euro (Dr. Reuse)
Leiter: W. Wahlster
Laufzeit: 2004-2007
LME
Chair for
Pattern Recognition
FAU Erlangen-Nuremberg
IMS Institut für Maschinelle
Sprachverarbeitung,
Universität Stuttgart
LMU
IPSK
Ludwig-Maximilians-
Universität München
ontoprise
SEMANTICS FOR THE WEB
SIEMENS
SYMPALOG
Speech Technologies
T Systems
ISI
UNIVERSITÄT
DES
SAARLANDES

SmartWeb: Mobile Access to the Semantic Web



HyLaP

Hybrid Language Processing Technologies for a personal associative information access and management application

Günter Neumann & Hans Uszkoreit

BMBF funded DFKI project, ~1.5 Euro

Start: 2006

Assumptions

- Question answering is the most natural way of requesting information
- The management of personal digital content becomes a true challenge
- The exploitation of personal digital memory will change our lives
- Applications for authoring, browsing and commenting will converge
- Every document can become an interface to memory

Exercise 1

- Try the following:
  - Enter into a public search engine a Wh-question
  - Inspect the best 10 documents for the correct answer
  - Identify the most important problems you had in finding an answer

Exercise 2

- Investigate the differences between IR, IE, QA
  - Where do they differ mostly ?
  - Why this aspects so specific ?
- Investiaget the commonalities between IR, IE, QA
  - What to they have on common ?
  - Where are possible interactions/interfaces ?

Exercise 3

- Textual question answering
  - Tries to find answers in text documents
- Structured question answering
  - Tries to find answers in data-bases
- Would it be possible to use the same approach in both cases or should one better define completely different architectures ?

Exercise 4

- Assuming two kind of document sources
  - A small (some GB) corpus of newspaper articles
  - The whole Web
- What are the major differences for identifying and extracting answer candidates ?