Question Answering

Günter Neumann

Language Technology Lab at DFKI

Saarbrücken, Germany
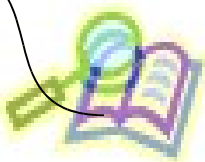
German Research Center for Artificial Intelligence

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

Experience
d-based
QA cycles

Search Engines

Answer Engines
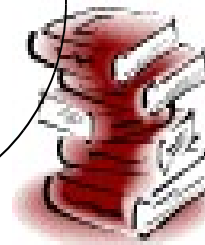
Shift more „interpretation
effort" to machines

User still carries the
major efforts in
understanding

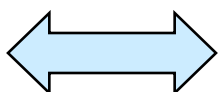German Research Center for Artificial Intelligence

ᵧ Input: a question in NL; a set of text and database resources

ᵧ Output: a set of possible answers drawn from the resources

"Where did Bill Gates go to college?"

"What is the rainiest place on Earth?"

QA SYSTEM

Text Corpora & RDBMS
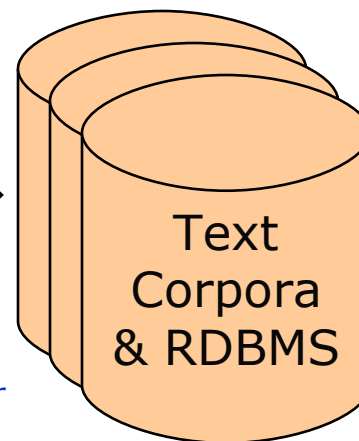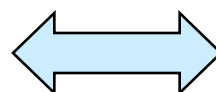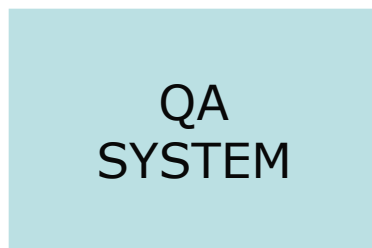
"Harvard"

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

"Mount Waialeale"

"… 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.)

German Research Center for

1

ɣ 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

- Open-domain questions & answers

- Information overload

  - How to find a needle in a haystack?

- Different styles of writing (newspaper, web, Wikipedia, PDF sources,…)

- Multilinguality

- Scalability & Adaptibility

German Research Center for Artificial Intelligence

''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)

- 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/ inconsistencies)

German Research Center for Artificial Intelligence

- Natural language questions, not queries

- Answers, not documents (containing possibly the answer)

- A resource to address 'information overload'?

- Most research so far has focused on fact-based questions:

  - "How tall is Mount Everest?",

  - "When did Columbus discover America?",

  - "Who was Grover Cleveland married to?".

- Current focus is towards complex questions

  - List, definition, temporally restricted, event-oriented, why-related, …

  - Contextual questions like "How far is it from here to the Cinestar?"

- Also 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?"

- Information Retrieval

  - Retrieve relevant documents from a set of keywords; search engines

- Information Extraction

  - Template filling from text (e.g. event detection); e.g. TIPSTER, MUC

- Relational QA

  - Translate question to relational DB query; e.g. LUNAR, FRED

ᵧ **Traditional QA Systems (TREC)**

– Question treated like keyword query

– Single answers, no understanding

Q: *Who is prime minister of India?*

<find a person name close to *prime*, *minister*, *India* (within 50 bytes)>

A: *John Smith is not prime minister of India*

- Future QA Systems
  - System understands questions
  - System understands answers and interprets which are most useful
  - System produces sophisticated answers (list, summarize, evaluate)

  What other airports are near Niletown?

  Where can helicopters land close to the embassy?

German Research Center for Artificial Intelligence

- Acquiring high-quality, high-coverage lexical resources

- Improving document retrieval

- Improving document understanding

- Expanding to multi-lingual corpora

- Flexible control structure
  - "beyond the pipeline"

- Answer Justification
  - Why should the user trust the answer?
  - Is there a better answer out there?

ꙟ Question: "When was Wendy's founded?"

ꙟ Passage candidate:

– "The renowned Murano glassmaking industry, on an island in the Venetian lagoon, has gone through several reincarnations since it was founded in 1291. Three exhibitions of 20th-century Murano glass are coming up in New York. By Wendy Moonan."

ꙟ Answer: 20th Century

- **Q336:** *When was Microsoft established?*

- **Diff cult** because Microsoft tends to establish lots of things…

  Microsoft plans to <u>establish</u> manufacturing partnerships in Brazil and Mexico in <u>May</u>.

- Need to be able to detect sentences in which `Microsoft' is **object** of `establish' or close synonym.

- Matching sentence:

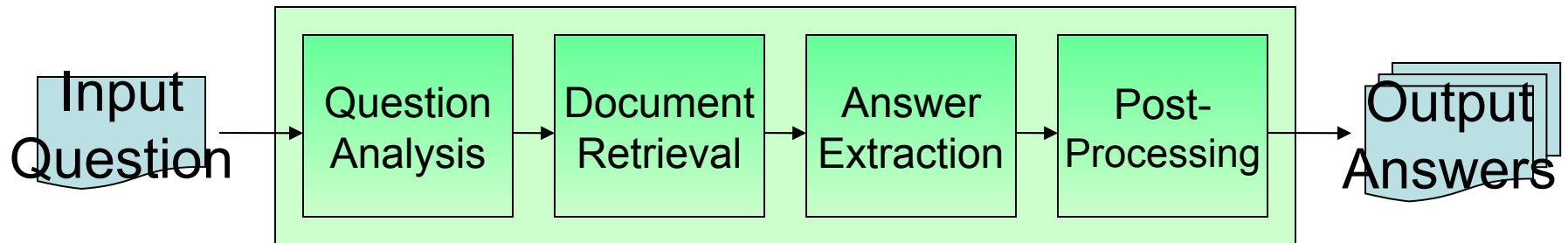 *Microsoft Corp was founded in the US in 1975, incorporated in 1981, and established in the UK in 1982.*

- Question: *What is the occupation of Bill Clinton's wife?*

    – No documents contain these keywords plus the answer

- Strategy: decompose into two questions:

    – *Who is Bill Clinton's wife?* = X

    – *What is the occupation of* X?

German Research Center for Artificial Intelligence

- The focus in the beginning of QA research was on closed-domain QA for different applications:

    - 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)

- At late 90th the focus shifted towards open-domain QA

    - TREC's QA track (began in 1999)

    - Clef crosslingual QA track (since 2003)

- ℞ Open domain

  - No restrictions on the domain and type of question

  - No restrictions on style and size of document source

- ℞ Combines

  - Information retrieval, Information extraction

  - Text mining, Computational Linguistics

  - Semantic Web, Artificial Intelligence

- ℞ Cross-lingual ODQA

  - Express query in language X

  - Answer from documents in language Y

  - Eventually translate answer in Y to X

German Research Center for Artificial Intelligence

Input Question → Question Analysis → Document Retrieval → Answer Extraction → Post-Processing → Output Answers

ʏ A sequence of discrete modules cascaded such that the output of the previous module is the input to the next module.

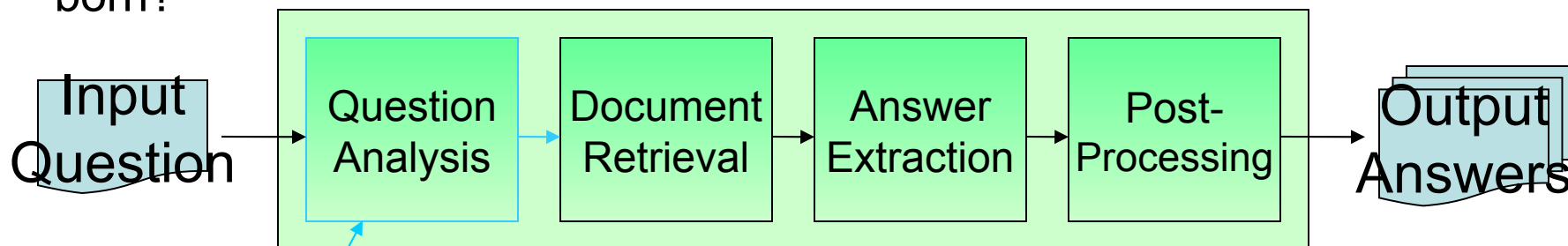German Research Center for Artificial Intelligence

"Where was Andy Warhol born?

Input Question → Question Analysis → Document Retrieval → Answer Extraction → Post-Processing → Output Answers

German Research Center for Artificial Intelligence

"Where was Andy Warhol born?

Input Question → **Question Analysis** → **Document Retrieval** → **Answer Extraction** → **Post-Processing** → Output Answers
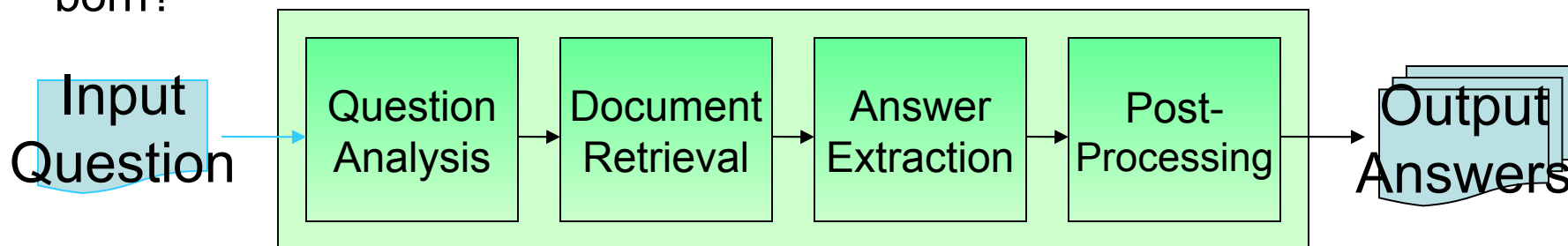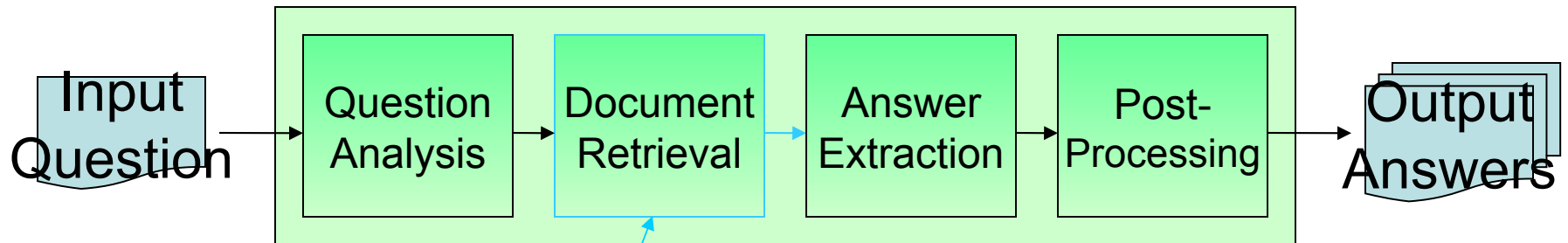
Discover keywords in the question, generate alternations, and determine answer type.

Keywords: Andy (Andrew), Warhol, born
Answer type: Location (City)

German Research Center for Artificial Intelligence

Input Question → **Question Analysis** → **Document Retrieval** → **Answer Extraction** → **Post-Processing** → **Output Answers**

Formulate IR queries using the keywords, and retrieve answer-bearing documents

(Andy OR Andrew) AND Warhol AND born

Input Question → Question Analysis → Document Retrieval → Answer Extraction → Post-Processing → Output Answers

Extract answers of the expected type from retrieved documents.
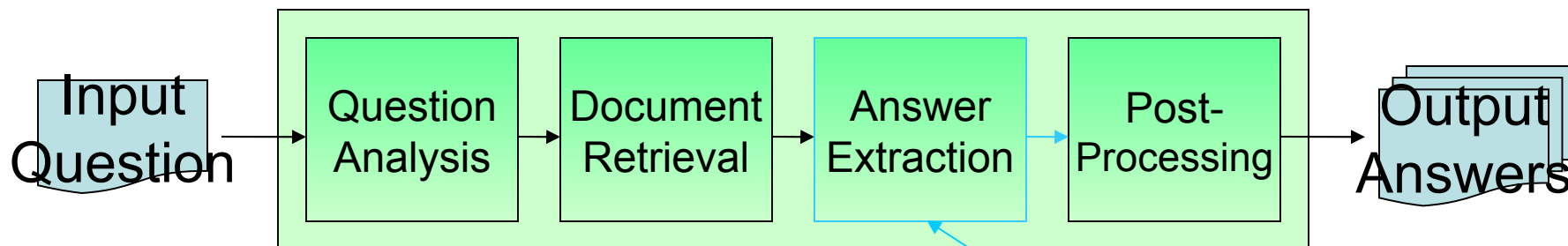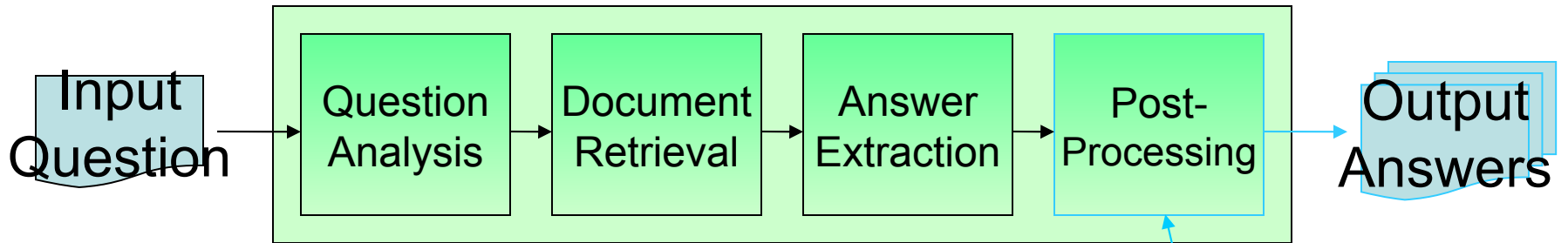
"Andy Warhol was born on August 6, 1928 in Pittsburgh and died February 22, 1927 in New York."

"Andy Warhol was born to Slovak immigrants as Andrew Warhola on August 6, 1928, on 73 Orr Street in Soho, Pittsburgh, Pennsylvania."

Input Question → Question Analysis → Document Retrieval → Answer Extraction → Post-Processing → Output Answers

Pittsburgh

merge

select appropriate granularity

1. 73 Orr Street in Soho, Pittsburgh, Pennsylvania

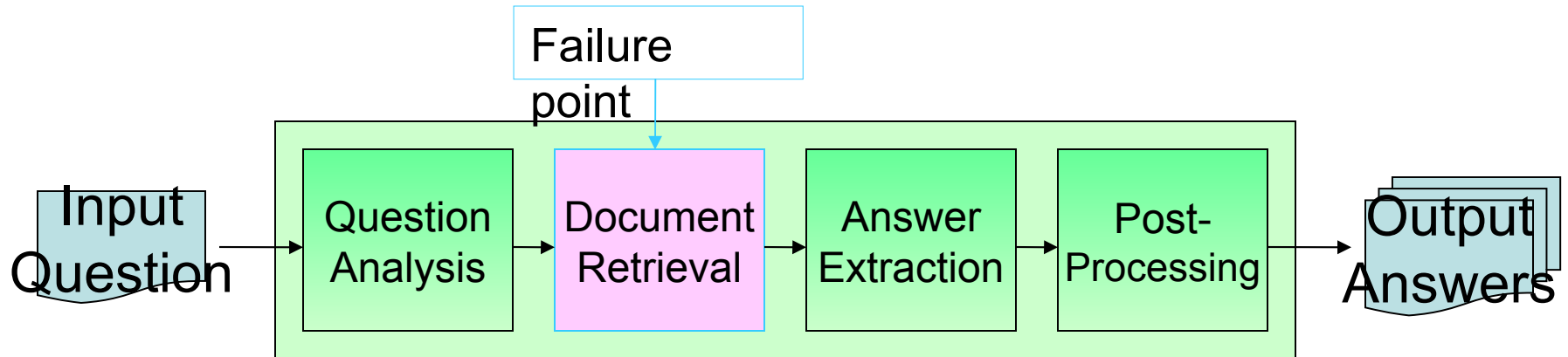2. New York

rank

Pittsburgh, Pennsylvania

1. "Pittsburgh, Pennsylvania"
2. "New York"

Answer cleanup and merging, consistency or constraint checking, answer selection and presentation.

Failure point

| Input Question | → | Question Analysis | → | Document Retrieval | → | Answer Extraction | → | Post-Processing | → | Output Answers |

- A pipelined QA system is only as good as its weakest module

- Poor retrieval and/or query formulation can result in low ranks for answer-bearing documents, or no answer-bearing documents retrieved

German Research Center for Artificial Intelligence

- ⲅ 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

- ⲅ 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

  - – Homepage: http://trec.nist.gov/data/qamain.html

- QA Track first introduced at TREC 8 (Voorhees, 1999)

  - 200 fact-based short-answer questions

  - Questions mainly back formulated from documents

  - Answers could be 50-byte or 250-bytes snippets

  - 5 answers could be returned for each question

  - Best systems could answer over 2/3 of the questions (Moldovan et al., 1999; Srihari and Li, 1999).

- TREC 10 (Voorhees, 2001) introduced:

  - List questions such as *"Name 20 countries that produce coffee"*

    - *Best 3 systems: 0.76%, 0.45%, 0.34% average accuracy* (computed as the number of distinct instances divided by the target number instances)

    - *Average for all 9 systems: 0.33 %*

  - Questions which don't have an answer in the collection (NIL answers)

- In TREC 11 (Voorhees, 2002):

  – Answers had to be exact

  – Only one answer could be returned per question

  – Best 3 systems: 83%, 58%, 54.2%, accuracy on 500 questions

  – Next systems: 38.4%, 36.8%, 35,8%, 28.4%, …

- TREC 12 (Voorhees, 2003) Introduced definition questions:

  – Define a target such as "aspirin" or "Aaron Copland"

  – A definition should contain a number of important facts (vital nuggets)

  – Can also include other associated information (non-vital nuggets)

  – Evaluated using a length based precision metric which penalizes long answers containing few nuggets.

    • Performance for the best systems: 0.555, 0.473, 0.461, 0.442, 0.338, 0.318

  – Final scores (fact, list, def questions) for best systems:

    • 0.559, 0479, 0.363, 0.313, 0.266, 0.256

German Research Center for Artificial Intelligence

- TREC 13 (Voorhees, 2004) combines the three question types into scenarios around targets. For instance

    – Target: Hale Bopp Comet

    – Factoid: When was the comet discovered?

    – Factoid: How often does it approach the earth?

    – List: In what countries was the comet visible on it's last return?

    – Other: Tell me anything else not covered by the above questions

- Performance of best systems:

    – 0.601, 0.545, 0.386, 0.278

German Research Center for Artificial Intelligence

- ⋎ Questions were based around 75 targets

  – 19 people

  – 19 organizations

  – 19 things

  – 18 events

- ⋎ The series of targets contained a total of:

  – 362 factoid questions

  – 93 list questions

  – 75 (one per target) other questions

- ⋎ All answers had to be with reference to a document in the AQUAINT collection of newswire texts.

- AMWAY

  - F: When was AMWAY founded?

  - F: Where is it headquartered?

  - F: Who is president of the company

  - L: Name the officials of the company

  - F: What is the name "AMWAY" short for?

  - O:

- return of Hong Kong to Chinese sovereignty

  - F: What is Hong Kong's population?

  - F: When was Hong Kong returned to Chinese sovereignty?

  - F: Who was the Chinese President at the time of the return?

  - F: Who was the British Foreign Secretary at the time?

  - L: What other countries formally congratulated China on the return?

  - O:

ᴦ Shiite

- F: Who was the first Imam of the Shiite sect of Islam?

- F: Where is his tomb?

- F: What was this person's relationship to the Prophet Mohammad?

- F: Who was the third Imam of Shiite Muslims?

- F: When did he die?

- F: What portion of Muslims are Shiite?

- L: What Shiite leaders were killed in Pakistan?

- O:

German Research Center for Artificial Intelligence