Off-line Strategies for Answering Dutch Questions

Maarten de Rijke

University of Amsterdam

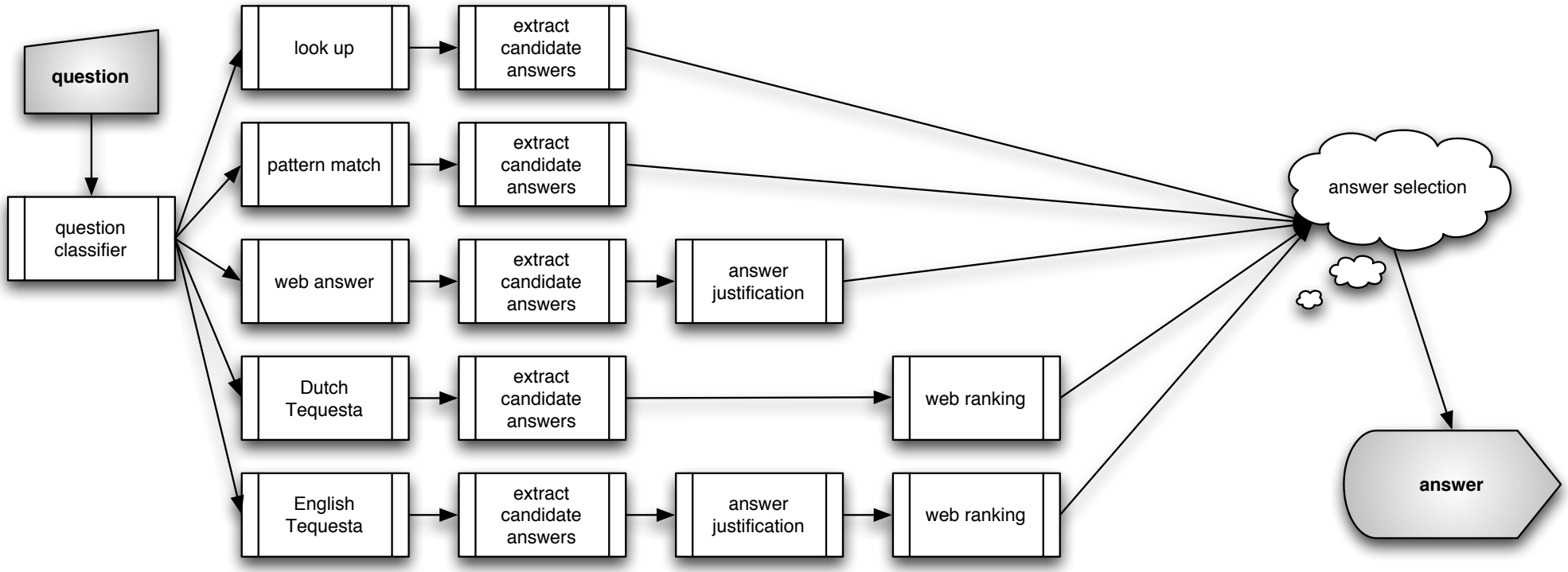
Cross Language Evaluation Forum 2003
Trondheim, Norway
August 22, 2003

Research funded by the Netherlands Organization for Scientific Research (NWO) under project numbers 220-80-001, 612-13-001, 365-20-005, 612.069.006, 612.000.106, and 612.000.207

Overview

- Joint work with Valentin Jijkoun and Gilad Mishne
- Have done question answering at TREC since 2001
  - the **Tequesta** system, a linguistically informed QA system for English
- Main interest
  - porting our English QA work to Dutch
  - creating Dutch resources
  - exploring multiple strategies
    - different anwering strategies required for different question types
    - e.g., abbreviation questions vs Who-did-what-to-whom
- Main challenge for Dutch
  - fewer resources and off-the-shelf tools (NE tagger, WordNet, . . .)

Architecture



Architecture, cont'd

- Five streams, each a complete QA system in itself
  - **Table Lookup**: use pre-constructed specialized knowledge bases
  - **Pattern Match**: search for answer patterns generated from a question
  - **Web Answer**: transform questions to Web queries, use Google to retrieve snippets of relevant documents from the Web and retain suitable phrases that occur significantly often
  - **Dutch Tequesta**: Tequesta with the language-specific components changed to Dutch (named entity and part-of-speech tagging, lemmatization, usage of WordNet, etc.)
  - **English Tequesta**: translate questions into English and use Tequesta (our TREC-2002 system)

Architecture, cont'd

- Many shared components
  - **Answer Justification**: given an answer found outside the test corpus, find a supporting document in the collection by means of information retrieval techniques
  - **Web Ranking**: use Web hit counts to re-rank and normalize confidence values for answers from different streams
  - **Answer Filtering**: filter out strings that are unlikely to answer the question, remove "noise" around answer strings, and merge similar answers
  - **Answer Selection**: select the best answer candidates, based on question type and confidence values provided by the streams
  - **NE tagger**: statistical n-gram tagger trained on the Spoken Dutch Corpus (CGN) combined with a tagger based on regular expressions
  - ...

Zoom in on the Table Stream

- Mine the text collection (off-line) to extract specific types of information
  - frequent TREC question types (translated in Dutch)
  - answers have a more or less regular shape (in Dutch)
  - abbreviations, capitals, currencies, inhabitants, leaders, locations, roles
  - perform a small amount of noise reduction
- Cascaded lookup mechanism to find relevant information

An Extract from the *Locations* Table.

Location 1	Location 2	Justification	Frequency
Bariloche	in Argentinie	NH19951016-0048	4
Barrow County	in Verenigde Staten	NH19950524-0117	2
Baskenland	in Spanje	NH19951130-0117	2
Basra	in Irak	AD19940104-0065	1
Basra	slechts vier kilometer van de grens met Iran	AD19941015-0007	1
Bathmen	bij Deventer	AD19940708-0171	1
Batna	350 kilometer ten oosten van Algiers	NH19940314-0029	1
Baucau	ten oosten van Dili	AD19950110-0064	1

A Youthful QA System

- **Q60**: Wie heeft de Berlijnse Muur gebouwd?
  (English: Who built the Berlin Wall?)
  - Answers: Afrikanen (English: Africans)
    frogs (English: frogs)
- **Q13**: Waar ligt Basra?
  (English: Where is Basra located?)
  - Answers: in Irak (English: in Iraq)
    slechts vier kilometer van de grens met Iran
    (English: only four kilometers from the border with Iran)
- **Q104**: Who is the president of Peru?
  (English: Who is the president of Peru?)
  - Answers: Alberto Fujimori
    Guerra NEDERLANDSE VERTALING

Results

- Ran 200 factoid questions against the Dutch CLEF corpus
- Special focus on the impact/success of the *Table Lookup* stream
- Results after a one hour bug fix of the CLEF 2003 system

Lenient Evaluation Results of CLEF 2003 Question Set.			
# Questions	Only *Table Lookup*	Without *Table Lookup*	All Five Streams
200 (all)	54 (27%)	64 (32%)	89 (45%)
187 (with answer)	41 (22%)	51 (27%)	76 (41%)

Conclusions

- Building on our TREC experience, we built a multi-stream QA for Dutch
- All streams contribute to the performance of the system
- The *Table Lookup* stream made a statistically significant difference in the final results, providing correct answers for 58% of all questions relevant to the stream
- Preliminary experiments suggest that weighted voting between the streams can further improve the overall preformance of the system
- During 2003 two new QA projects got funding in the Netherlands, which should lead to three participating teams for Dutch in QA@CLEF