Question Answering
What's Next?

Maarten de Rijke
Informatics Institute
University of Amsterdam
Focused Retrieval

- What is it?
- Why?
- How?
Information

● Avalanche
  ● Internet
  ● Intelligence
  ● Science (astronomy, biomedical, …)
  ● Desktop

● Growing at an growing pace
  ● 1999: 250 MB pp for each person on earth
  ● 2002: 800 MB pp for each person on earth
Challenges

- Pinpointing
  - Highly relevant, highly specific retrieval
- Retrieving not just documents
  - People, objects, places, services, answers, …
- Multiple languages, tasks, information needs, types of users
Facing the challenge

- Language understanding?
- Method
  - Mix theory, experiment, application
- Task-driven
  - Real users, real information needs, real data, or at least “approximations”
QA as pinpointing

• Given a document collection and a question, return the answer
• Attractive task
  • Opportunities for language processing in information access?
  • Opportunities for knowledge representation in information access?
  • Combine IR, NLP, AI
• People ask questions!
People ask questions

Search engine log file
Sep 22, 2004–Jan 9, 2005
10M searches

99%
1%

Queries
Questions

Distribution of questions from the log

48%
26%
5%
4%
4%
3%
9%

How
When
Why
Other
What
Where
Who
QA history

- Around since the 1960s
  - Natural language front-end for DBs
  - LUNAR

- Focus on open-domain since 1990s
  - MURAX (Kupiec 1993)
  - Hirschman: reading comprehension
  - TREC QA: 1999–…
TREC

- Text REtrieval Conference
  - Launched in 1992
  - Large general full text files
  - Information analyst needs
  - Zillions of tasks (over the years)
    - Ad-hoc retrieval, routing/filtering, cross-language, scanned documents, spoken documents, video, NLP, web, question answering, novelty, robust, genomics, HARD, terabyte, enterprise
QA at TREC

- Launched at TREC-8, 1999
- Recent years
  - Focus on factoids ("NE retrieval")
    - Single exact answer with supporting doc
  - Variations on factoids ("lists")
  - Definitions (2003)
    - Return “nuggets”
  - “Other” questions (2004)
Factoid examples

- Factoids
  - Who discovered Oxygen?
  - When did Hawaii become a state?
  - Where is Ayers Rock located?
  - What team won the World Series in 1992?

- Lists
  - List the names of chewing gums
  - What governments still officially recognize and support International Labor Day?
Definitions/”Others”

● Definition
  ● Who is Aaron Copland?
  ● What is a quasar?

● “Other” organized around a target
  ● Target: Americorps
  ● Factoid: How many volunteers work for it?
  ● List: What activities are its volunteers involved in?
  ● Other: Tell me other interesting things about this target I didn’t know enough to ask directly?
QA dimensions

- Nature of the information
  - Structured, semistructured, unstructured
- Nature of the questions
  - Factoids, definitionoids, procedures, opinions
- Nature of the answer
  - Extracted snippet, generated text
- Nature of the technique
  - Linguistically sophisticated, data-driven
So?

Let’s step back
TREC QA ingredients

a. Questions: general purpose
b. Data sources: open domain
c. Answers: single, one shot, no follow-up
d. Users/user models/…
Issues

- What’s the nature of the user’s need?
- What’s the appropriate response?
- What’s the appropriate data source?
Assumptions …

- ... about questions
  - Forget the information analyst (who comes with a rich context, set of tools, scenario, task, ..., none of which is modeled in the TREC setting)
  - “Ad hoc” setting: any topic, any user, general purpose functionality
What do people ask?

- Back to the log files...
- Respectable amount of TREC factoids—not the majority

- how tall is christina aquilera
- what does mig welding stand for
- where does sean hannity live?
- how long does ritalin stay in your bloodstream?
- where does moss grow
What do people ask?

- Back to the log files...
- Respectable amount of TREC factoids—not the majority
- Lots of procedural questions
What do people ask?

- Back to the log files...
- Respectable amount of TREC factoids—not the majority
- Lots of procedural questions
- Large amount of definitions
What do people ask?

- Back to the log files...
- Respectable amount of TREC factoids—not the majority
- Lots of procedural questions
- Large amount of definitions
- Gems, many gems
- How tall is Christina Aguilera
- How to repair scratches in leather
- What is pink noise
- What is a catapult
- What is a pictograph
- Define social justice
- What is a rational number?
- How to understand women
- How to stop your dog from pooping on your stairs
- Almost everyone sees me without noticing me, for what is beyond is what he or she seeks. What am I?
- How smart are blonds
Recommendation

- If you’re interested in being able to answer questions that Real People ask
  - Do factoids
  - Do definitionoids
  - Do procedural questions
Assumptions …

• … about answers
  • Answers need not match to be helpful
  • Users apply their own selection
  • Answers are not end-points, but stages in a berry-picking process
Example

- Where is the Rijksmuseum located?
  - In a building designed by Cuijpers
  - Across Museumplein from the Concertgebouw
  - You can take Tram 5 to get there
  - Jan Luijkenstraat in Amsterdam
  - The Rijksmuseum van Oudheden (RMO) is the national museum of antiquities at Leiden

- What is the right answer?
Value of unsought info

- What’s the average rent for an apartment in Amsterdam?
  - Try this site: http://www.expatriates.com/classifieds/amst/
- Who was Theo van Gogh?
  - Van Gogh made the movie ‘Submission’ in collaboration with Ayaan Hirsi Ali, a Dutch politician and former Somali refugee
Recommendations

- Return ranked lists of answers
- Answers in context so that users (assessors) can determine validity
- Continue with “other” questions
- Work on novelty and importance models
Assumptions …

- … about documents
  - Whatever we define our user or user model to be, the data sources should be “appropriate” for the intended questions
  - Huh?
What data source?

- To look up an angioprim?
- To find out how fast cheetahs runs
- To find out about UK economic growth?
- To determine what Mozart was born?
- To repair a cold air leak in Mazda?
Data sources

- The web
- Or heterogeneous
  - dictionary for definitions
  - manuals/FAQs for procedural
  - encyclopedia/scientific papers/… for general knowledge questions
  - newspapers for current events
Recommendations

• Avoid document-question mismatch
• Do Web QA
• Or do QA against a heterogenous collection, where resource selection is a key part of the task
Question Answering
What’s Next?
QA evaluations in 2005

- TREC
  - Pretty much the same as in 2004
- CLEF
  - Debating scenario, doc collection, type of questions, answer format, …
- INEX
  - …
INEX

• What is INEX?
  • INitiative for the Evaluation of XML retrieval
  • Launched in 2002
  • Two types of information needs
    • CO: content only
    • CAS: content and structure
  • IEEE computer science journals
  • Provide focused access by returning elements, not documents
  • $1M question: Does structure help?
QA at INEX

- Organized around a user model
  - Undergraduate student writing essays
- Questions
  - Factoids, definitions, procedural, …
- Documents
  - Semistructured, scientific articles, encyclopedia, dictionaries
- Answers high-lighted in context
QA at UAmS
What’s Next?
QA at UAmS

- Participation
  - TREC, CLEF, INEX
- Coordination
  - INEX, Dutch QA for CLEF
- Research
  - Multi-stream architecture for QA
  - …
QA research at UAms

- ... 
- NERC and IE for QA (emphasis on Dutch) 
- QA against Wikipedia and similar resources 
- FAQ mining and retrieval 
- Developing novelty and importance models 
- ... 
- Contemplating to open source a “dressed down” QA system
Question Answering
What’s Next?

- Lots
- And it will be great
- And you should be part of it