Statistical Query Transformations for Question Answering in the Web

Ilya Boyandin
St. Petersburg State University, 2003

Question Answering

Goal:

Find a short text fragment which answers the question.

Example:

***Q**: Who invented the light bulb?*

***A**: Thomas Edison*

Question Answering

- Retrieve answers rather than documents.
- Precision is important, recall isn't.
- The "correct" answer is one that can be found in the collection.
- No new knowledge is produced.

TREC: Text REtrieval Conference

TREC QA Track

- Collection of documents (newspaper)
- Test set of questions

TREC-9 (2002)

- 979,000 documents (3Gb of text)
- 682 questions (Encarta log, Excite log)
- Best resut: 65% of questions answered (Falcon)

Using Web for QA

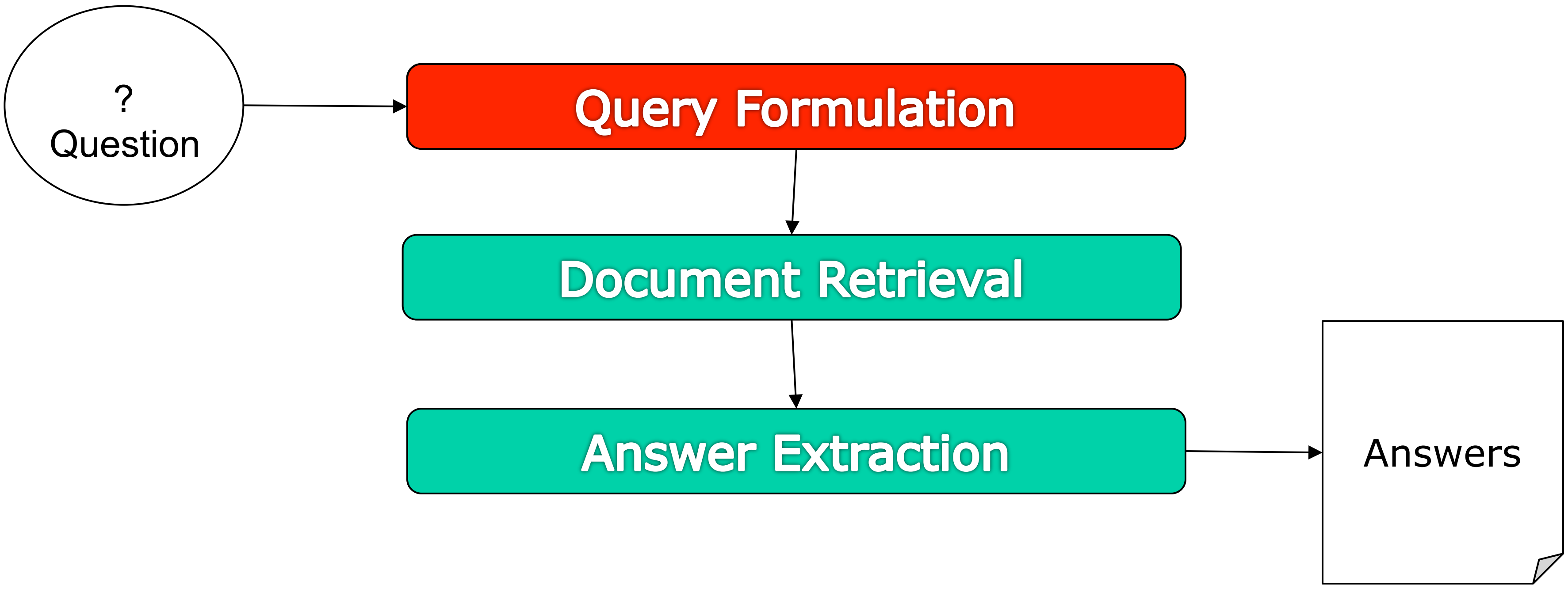
Benefits

- Vast number of answers
- Constant updates
- Redundancy

Challenges

- Wrong and contradictory answers
- Undated information
- Heterogeneous, irregular structure

Components of a QA System



Query Formulation

Goals:

– “Translate” the question into the form the IR engine understands

– Narrow the set of documents to consider

Example:

Question*: When was Nabokov born?*

Query*: Nabokov /4 !born*

Question Transformation

Types of Transformations:

- Remove question words/other words
- Add words/phrases likely to be in the answer
- Add synonyms
- Morphological changes
- Add query language operators
- ...

Why Learning?

- Hard to predict what transformations will be better
- Far from all patterns are obvious

Evaluating Transformations

TRDR Metric:

– Query is sent to the IR engine

– Positions of the documents containing the right answers: r_i (of the first N returned documents)

$$TRDR = \sum_{i=1}^{n_{correct}} r_i^{-1}$$

Example:

$$TRDR = \frac{1}{2} + \frac{1}{4}$$

Radev et al 2001

QASM Algorithm

Atomic Operators

- E.g. add/remove/substitute words
- Transformation is a composition of atomic operators

Query Features

- E.g. type, number of words, number of nouns
- Context of a query: values of all features
- Questions/queries with the same context are treated in the same way

Radev et al 2002

QASM: Learning

Training Set

Questions with answers

Iterative Learning

1. Applies every atomic operator to the query
2. Submits to the IR engine
3. Evaluates results, updates the model
4. Applies the best operator to the query
5. Next iteration

Resulting Model

Allows to find the best (statistically) operators for any context

Radev et al 2002

QASM: Question Transformation

Input:

Question

Iterative Transformation

1. Calculates the context of the query
2. Finds the best operator for the context
3. If it's IDENTITY then stops
4. Applies the best operator to the query
5. Next iteration

Output:

IR engine query

Experimental Environment

Test set

100 questions from the Yandex log

Atomic Operators

- Remove words (based on frequency)
- Add query language operators: distance; restrict morphological changes

Query Features

Question type; Number of words; Number of nouns

QASM Analysis

Results

No quality improvement in most cases

Problem

The selectivity of the generated transformations was often too low or too high

Possible Causes

- Too small training set
- ~~Choice of operators/features~~
- Irregularity

Who won the Nobel peace prize in 1975?
Who won the Nobel peace prize in 1979?

"Optimal" QASM

- Same atomic operators and query features
- Use only the best transformation for each question
- Improvement ~50%

mQASM

Changes to QASM

- Generates a set of best queries ordered by selectivity
- Submits queries until there is enough results
- Weights results and builds one ordered list

Evaluation Results

Stability of the Comparisons

- 40 random combinations

  60 questions for training, 40 for testing

- Significance level: 5%
- Wins/Losses/Draws

	Yandex			QASM		
QASM	2	29	9	-		
mQASM	37	0	3	40	0	0

Evaluation Results

Same Experiment with Google

Environment is the same as with Yandex, except using less atomic operators

	Google			QASM		
Max	15	0	25			
QASM	12	4	24	-		
mQASM	15	3	22	8	4	28