Answering Learners' Questions by Retrieving Question Paraphrases from Social Q&A Sites

Delphine Bernhard & Iryna Gurevych
Ubiquitous Knowledge Processing Lab, Darmstadt, Germany

Workshop on Innovative Use of NLP for Building Educational Applications, ACL 2008
June 19, 2008
Motivation: Information Overload

Which algorithms compute the shortest path in graphs?

Web contents: web pages, wikipedia, e-books

Course material: presentations, course exercises

Social learning: forums, blogs, mailing lists, chats, wikis
Motivation: Information Overload

Which algorithms compute the shortest path in graphs?

The most important algorithms to solve this problem are:
- Dijkstra's algorithm
- Bellman Ford algorithm
- A* search algorithm
- Floyd-Warshall algorithm
- Johnson's algorithm
- Perturbation theory

Web contents:
- web pages
- wikipedia
- e-books

Course material:
- presentations,
- course exercises

Question-answering system

Social learning:
- discussion forums,
- mailing lists,
- chats, wikis, blogs
Conventional QA Systems

- Challenges in the educational domain:
  - Open-ended, explanatory or methodological questions
  - Quality of both the questions and the answers
Where to start?

• Many questions have already been asked by somebody else before
• Idea: Find already answered questions in discussion forums, FAQs, Social Q&A sites, etc.
• Problem: find an answered question which is identical in meaning to the question asked by the user

Question Answering as question paraphrase identification in Q&A repositories
What are question paraphrases?

- Question paraphrases are questions which:
  a) have identical meanings
  b) have the same answers
  c) present alternate wordings

- Example:
  - How many ounces are there in a pound?
  - What's the number of ounces per pound?
  - How many oz. in a lb.?
Question and Answer repositories

- **FAQs**
  - Questions and answers are compiled and subject to editorial control
  - Example: [www.faqs.org](http://www.faqs.org)

- **Ask-an-expert Services**
  - Provide expert answers to user questions
  - Examples: [www.madsci.org](http://www.madsci.org), [www.hhmi.org](http://www.hhmi.org)

- **Social Q&A sites**
  - Provide portals where users can ask their own questions and answer questions from other users
  - Examples: Yahoo! Answers, WikiAnswers
What is the weight of a polar bear?

For example: What is the secret of true love? | What is the best operating system and why?

Other contributors have said "What is the weight of a polar bear?" is the same question as "What is the average weight of a polar bear?". If you believe that these are not asking the same thing and should be answered differently, click here.

Q: What is the average weight of a polar bear?

Answer

Adult Males: 900-1500 lbs, avg = 1150; record = 2210 lbs. Males are 25-45% larger than Adult Females who weigh 330-560 lbs.
Question Paraphrase Dataset

- Extracted from WikiAnswers

- User annotated paraphrases
  - "What is the average lifespan of a polar bear?"

Our contributors said this page should be displayed for the questions below. *(Where do these come from 🤔)*

If any of these are *not* a genuine rephrasing of the question, please help out and edit these alternates.

Polar bear lifespan? Average age of a bear? A polar bear’s lifespan? How long do polar bears live? What is a Polar Bears lifespan? What is the polar bear lifespan? Average lifespan of a polar bear? Average lifespan for a polar bear? The Average Lifespan of a Polar Bear? What is the lifespan of a polar bear? What is a polar bears average lifespan? What is the average lifespan of a koala bear? What is the average life span of a polar bear?
Question Paraphrase Dataset

- 1,000 **target questions** with their paraphrases
- From 100 randomly selected files in the Education category
- 7,434 paraphrases → **input questions**
- Task: retrieve the corresponding target question for each input question
Question Paraphrase Identification

Pre-processing:
- stemming
- lemmatisation
- spelling correction

Input question

Target question

WikiAnswers

Question similarity measures

String similarity:
- matching coefficient
- overlap coefficient
- edit distance
- ngram overlap

Vector space:
- term vector similarity
- Lucene
Results

- Vector-space based methods outperform string similarity
- Morphological pre-processing and spelling correction do not ameliorate the results
Stemming and Lemmatisation

- Positive effects
  - What are analogies for mitochondria?
  - What is an analogy for mitochondrion?

- Negative effects
  - How was [be] calculus started [start]?
  - How could you start your MA English studies?
  - Who developed calculus?
Spelling Correction

- Positive effects
  - How do you become an anesthesiologist?
  - How many years of medical school do you need to be an anesthesiologist?

- Negative effects
  - What are the GRE score required to get into top100 US universities?
  - What are the GRE score required to get into top100 US universities?
Errors in the user generated gold standard

- Target question:
  What is the secret about mona lisa?

- Reformulations in WikiAnswers:
  When was the first painting made?
  Where did leanardo da vinci live?
Challenges in Question Paraphrase Identification in Social Q&A Sites

- Vocabulary mismatch:
  - What events occurred in 1919?
  - What important events happened in 1919?
- Spelling errors
  - When was indoor plumbing invented?
- Planned improvements:
  - Semantic relatedness metrics
  - Reliable spelling error identification and correction, including real-word spelling errors
Next steps

- Human-annotated question paraphrase gold standard
  - Question paraphrases
  - Related questions, on the same topic
  - Unrelated questions
- Automatic assessment of the quality of the answers
  - Correctness, well-formedness, readability, objectivity, relevance, utility, interestingness (Jeon et al., 2006; Agichtein et al., 2008)
- Automatic assessment of the quality of the questions
  - Interactive question answering: help learners ask good questions (Harabagiu et al., 2005; Rus et al., 2007)
Summary

- Realistic solution to the problem of automatically answering learners' questions: use repositories of already answered questions

- Method: identification of question paraphrases

- Resource: new question paraphrase dataset extracted from a social Q&A site, WikiAnswers

- Findings: vector space based measures outperform string similarity measures, morphological pre-processing and spelling correction do not improve the results

- Challenges: question quality and pre-processing
Acknowledgments

Ubiquitous Knowledge Processing Lab
http://www.ukp.tu-darmstadt.de
Comparison of the methods

- Dissimilar error patterns for different pre-processing strategies and paraphrase identification methods
- Best combination: Lucene, Term Vector similarity with stemming, Ngram overlap with spelling correction
  88.3% accuracy (best without combination: 87.4%)