brmson (YodaQA)

A DeepQA-style Question Answering Pipeline

Petr Baudiš ⟨pasky@ucw.cz⟩

FEE CTU Prague; brmlab hackerspace

Fall 2014

brmson: Question Answering

A Question Answering system inspired by **IBM Watson**
and its DeepQA pipeline architecture.

Open Domain Factoid Questions:
What is the name of the southwesternmost tip of England?
Who received the Peace Nobel Prize in 2014?
What hair color did Thomas Jefferson have before grey?
What is the distance of Earth from Sun?
How to change a flat tire?

How It Works

- Question is analyzed, clues and lexical answer type extracted
- Fulltext search for clues (we use English Wikipedia)
- *Hundreds* of candidate answers are generated from matching passages, introduction passages and document titles
- Candidate answers are scored based on various features (logistic regression)
- Important features: Type coercion
  "Is the answer a color?"
  "Is the answer an inventor, or at least a person?"
- Top scored answer is yielded

Current State

Immediate aims: Open-domain factoid questions (TREC QA),
replicating the DeepQA scheme with target
75% recall, 35% accuracy-at-1.

Current performance: 70.0% recall, 22.5% accuracy-at-1.

Work in progress: Structured datasets, evidence gathering.

brmson: YodaQA Implementation

- **YodaQA:** "Yet anOther Deep Answering pipeline"
- Designed and implemented from scratch
- Java, UIMA framework
- Architecture based on simplified DeepQA (as published)
- NLP analysis: Third-party UIMA annotators via DKPro
- **Open Source!** Everything is on `github.com/brmson`, including documentation
- Looking for contributors, collaborators, commercial ideas…

Conclusion

- Practical, open source QA system
  - Clean architecture and development methodology
  - Reasonably documented!
  - Clear path forward, towards reference experimental testbed
  - Immediate tasks: Add evidence gathering,
    query structured data sources

<div align="center">

pasky@ucw.cz
petr.baudis@gmail.com

Thank you for your attention!

</div>