

# brmson (YodaQA)

A DeepQA-style Question Answering Pipeline

Petr Baudiš [⟨pasky@ucw.cz⟩](mailto:pasky@ucw.cz)

FEE CTU Prague; brmlab hackerspace

Spring 2015





# Question Answering?

## Open Domain Factoid Questions:

What is the name of the southwesternmost tip of England?

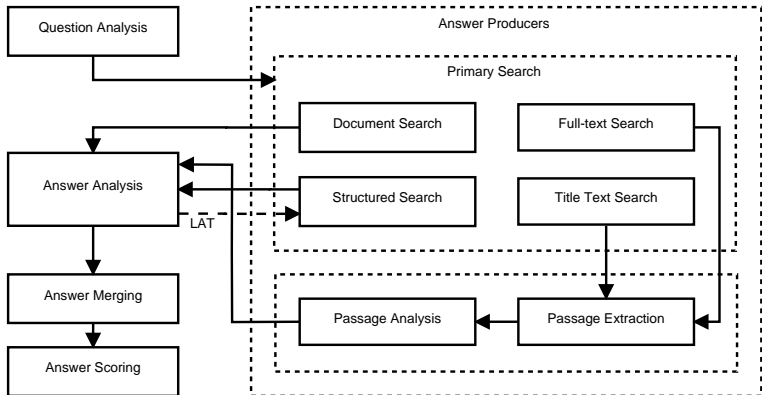
Who received the Peace Nobel Prize in 2014?

What color hair did Thomas Jefferson have before grey?

What is the distance of Earth from Sun?

How to change a flat tire?

# How It Works?



# 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



# brmson: YodaQA Implementation

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



# Conclusion

- Practical, open source QA system
- Clean architecture, very modular system
- Reasonably documented!
- Long term:
  - Closed domain QA with powerful user interface
  - Bleeding edge NLP research (PhD)
  - **Startup aims**

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

**Thank you for your attention!**