# broccoli

Emilie McConville
Zach Paine
Mattt Thompson
Owen Yamauchi

11-411 Natural Language Processing
Noah A Smith, Spring 2008

1

Hello, we're team Broccoli
(Like everyone else who couldn't come up with a team name, we got a vegetable, but as far as we're concerned, broccoli is the best)

# Approach
## (How we thought it should work)

From the start, we had some good ideas of how to approach the task at hand. However, if we were to summarize our plans into one concept, it would be...

Named Entity Recognition
Why?
- Most elements of interest are going to be uncommon words / entities
- Use this fact to generate interesting questions

# Also,

- POS Tagging
- Stemmer
- Anaphora /
  Co-reference Resolution
- Article Classification

We also anticipated other traditional NLP tools to be useful to refine what we were generating with our NER approach

# Implementation

## (How it actually ended up working)

...Of course, plans are subject to change, and as we went along with the project, we found out what worked and what didn't.

We found out that NER was too slow, and its results weren't that great. (right?)

# Keyword Matching

## (For Question Answering)

What we found effective:
For answering questions, we would remove common words (articles, auxiliaries) and collect remaining tokens. We would search through article and find the sentence with the most matches and return it as the answer.

# Keyword Matching

**Question:** *"Who was the sixteenth president of the United States?"*

**Keywords:** [sixteenth, president, united, states]

For answering questions, we would remove common words (articles, auxiliaries) and collect remaining tokens

# Keyword Matching
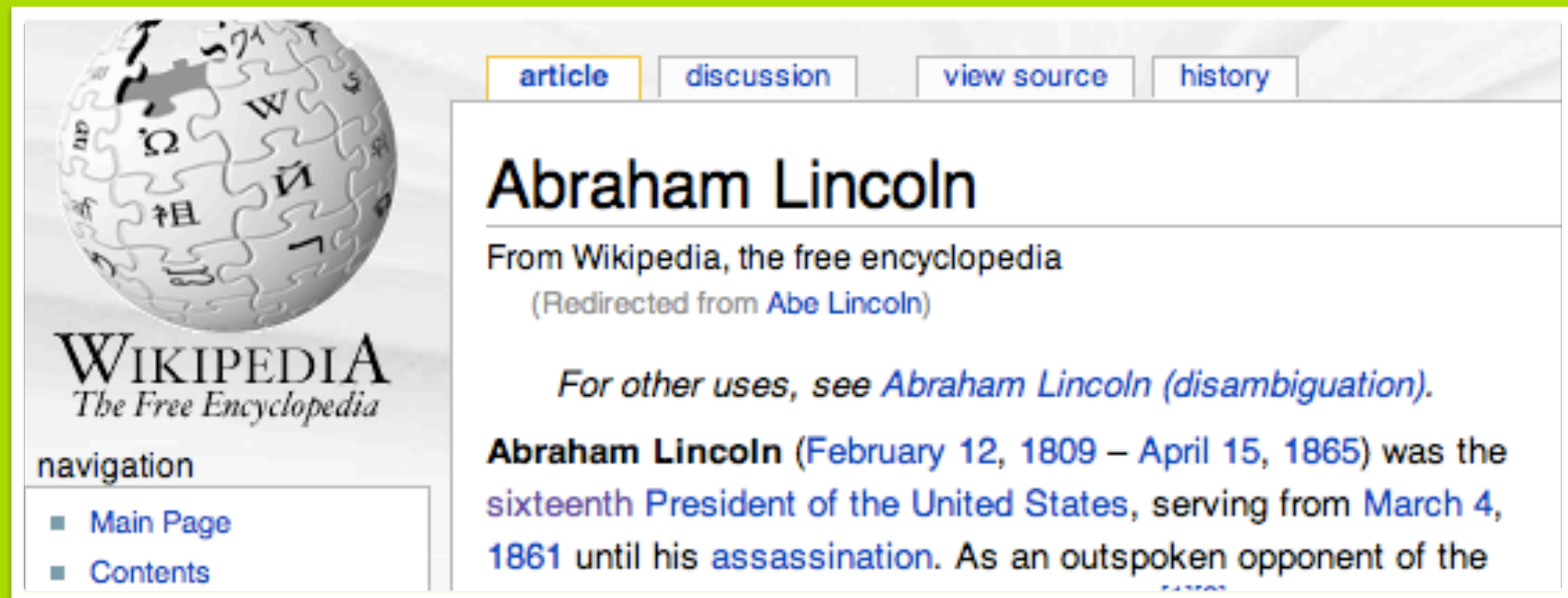


**Keywords:** [sixteenth, president, united, states, serving, assassination]

**Score:** 4/6 ✓

We would search through article and find the sentence with the most matches and return it as the answer

# Syntax Parsing

## (For Question Asking)

Using the Stanford Parser, we found sentences in the article that corresponded to sentence structures that we identified to be good indicators of a structure that we could turn into a question.

# Syntax Parsing

**Search With:** FactoidMatcher
DateMatcher

We would search through article and find the sentence with the most matches and return it as the answer

# Syntax Parsing
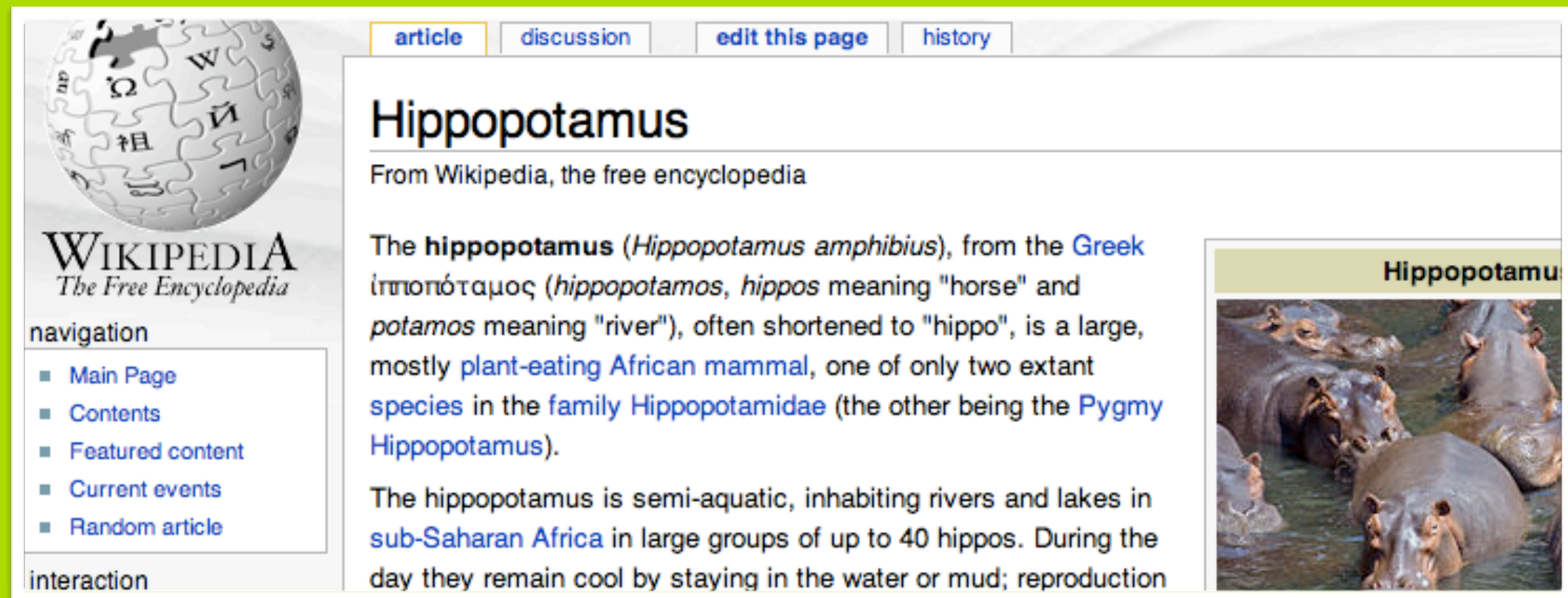


**Candidate:** *"They emerge at dusk to graze on grass"*

**Question Form:** "Who emerges at dusk to graze on grass"

After finding a candidate sentence, we use the POS tagging data to determine the correct question word and use a stemmer to conjugate the following verb correctly.

(Slide about unused/other possible strategies for next version: Classifier, Re-introduce NER, etc.)
(Any other ideas?)

(Slide about pragmatics: Ruby as DSL and glue language)

(Slide about pragmatics: SVN/ Basecamp/ Twitter)

# broccoli

Emilie McConville
Zach Paine
Mattt Thompson
Owen Yamauchi

11-411 Natural Language Processing
Noah A Smith, Spring 2008

Thanks!
Q/A