No GloVes Required

Rule-based Q/A System
**CS 6340 : Natural Language Processing**

Outline

- Overview
- System Architecture
- Components and Algorithms
- Performance
- Success and Regrets
- Lessons Learned
- Contributors

Overview

- Simple Rule-based Question Answering System
- Components-
    - **Information Retrieval System (IR) -** Returns ranked list of sentences based on scores given to words based on different parameters
    - **Information Extraction System (IE)**- Question Typing. Answer words extracted using rules based on question type.
    - **Selector System (SS)**- A feedback system which assesses the output returned by IE is fit to be the final answer
- SS allows the IE System validate and provide feedback to the IR System
- Use NLP tools namely Lexical Similarity, POS tagging, NER , Dependency Parsing

Peer System (PS)

System Architecture
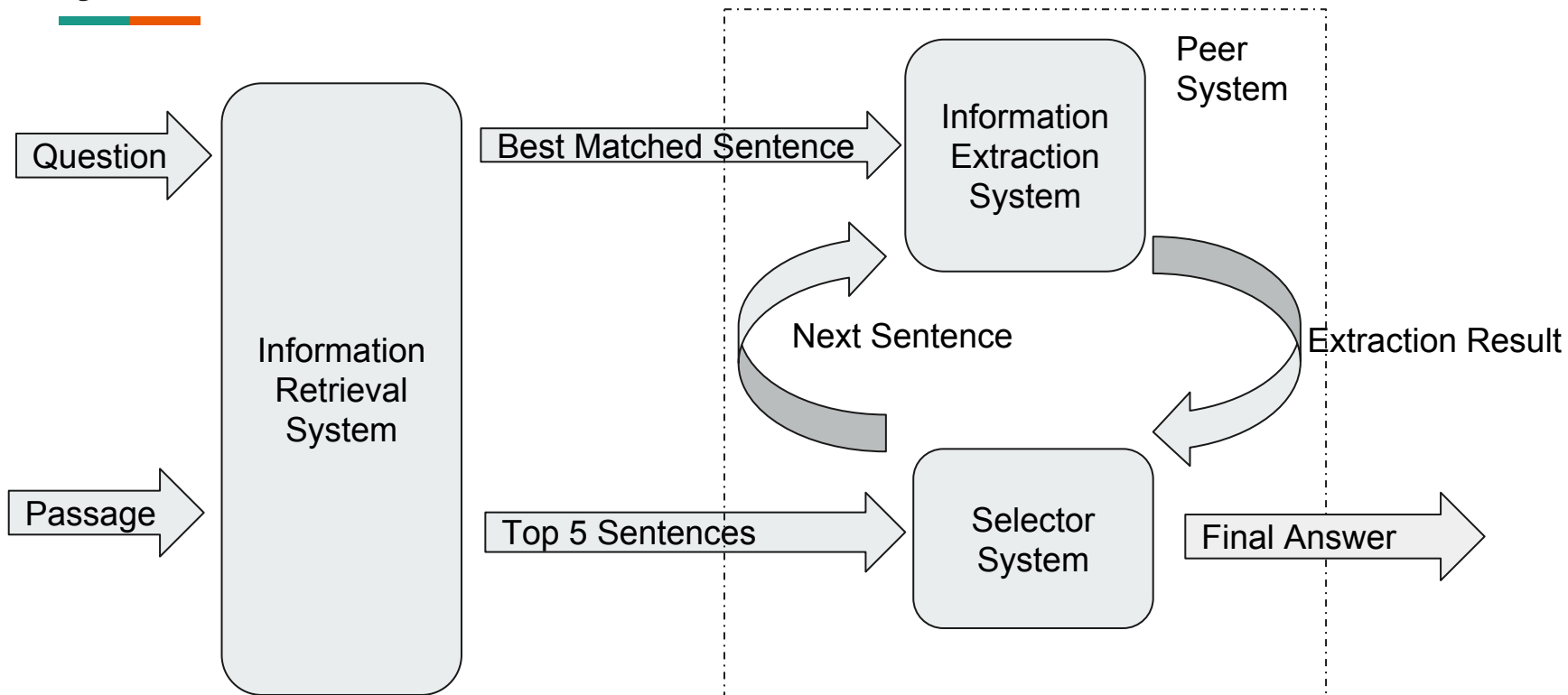
Information Retrieval System

- Each sentence is compared with question (word by word).

- Words are matched on multiple parameters [1] -

    - **Word Match Score *(wmsc)*** - Syntactic Similarity [2] based string matching

    - **Strike Score *(strksc)*** - Higher reward for consecutive matched tokens

    - **POS Score *(possc)*** - Certain POS tags match words get higher scores

    - **Root Score *(rootsc)*** - root and xcomp match words get higher scores

    - **Inverse of Term Frequency *(invtf)*** - weighted by inverse of word count

$$WordScore(w_i) = (wmsc + strksc + possc + rootsc) * invtf$$

$$SentenceScore(S_j) = \frac{\sum_{i=1}^{N} WordScore(w_i)}{N}$$

Information Extraction System

Extract information from retrieve sentence. Focuses on improving the answer recall.

1. Question Typing

2. Question type specific extraction rules [3] (NER, POS, and Dependency tags)
   a. **Who:** "PERSON", "ORG", "NORP", and "GPE"
   b. **Where:** "LOC" ,"FAC", "GPE", "PRODUCT", "WORK_OF_ART", "LAW", and "ORG"
   c. **When:** "DATE", and "TIME"
   d. **How much:** "MONEY", "QUANTITY', "CARDINAL", and "PERCENT" (if "time" : "TIME")
   e. **How many:** 'QUANTITY',"CARDINAL", and "PERCENT" (if "years, days" : "DATE")
      i. More similar rules on basis of dictionary of words
   f. **Which, What, How and Why:** Looks for 'Cluewords' in sentences
   g. **Does, Is, Was:** 'YES' or 'No' (handle negation by dictionary)

3. Smartly remove question words and punctuation characters to improve recall

Selector System

1. **Validate the output** of the Information Extraction (IE) System

   a. **Satisfactory :** Output extracted information

   b. **Otherwise :** Supply next best retrieved sentence to the IE system

      ■ Repeat 1

2. If nothing substantial extracted from the top 5 sentences return information extracted from the top rank sentence

Benefit : Handles the special case where complete or empty sentence is returned. Frequently happens in *Who, Where, and When* questions and few variants of *How* questions.

Performance

Question Type	Precision	Recall	F-Score	#Questions
Where	0.4180	0.4359	0.4268	46
Who	0.4330	0.377	0.4031	47
When	0.6826	0.6783	0.6804	34
What	0.2925	0.6502	0.4035	93
Why	0.4289	0.5479	0.4812	27
How much & How many	0.5871	0.6959	0.6385	41
How	0.4029	0.66	0.5003	17
Which	0.3533	0.3967	0.3737	3

Detailed QA Performance Sheet - https://goo.gl/HknUfQ

Success And Regrets

- **Success-**
    - **Our Peer System :** Significantly increase precision
    - **Our Sentence Scoring :** Penalize frequent words
    - **Our NER-IE Rules:** High recall on *{When, How much, How many}* questions
    - **Best absolute score** in precision and recall
- **Regrets-**
    - **Coreference resolution** fails to improve system performance
    - Doesn't take into account **semantic similarity** (e.g. Word Mover Distance)
    - Doesn't incorporate **pattern based category extraction** (e.g. groups of people)
    - Doesn't **automatically learn IR scoring weight**s (e.g. Linear programming**)**
    - Better extraction technique for *{What, Why, How}* **questions**

Lessons Learned

- **Best retrieved sentence** may not have the correct answer

- Room for improvement in **Coreference Resolution**

- **Verbs** act as the biggest clues in questions

- **Pattern extraction** based on question type is a tricky problem

- An algorithm for **automatic learning of score weights** is desirable

- **Preprocessing** is not as trivial as you think it is!

CONTRIBUTORS

MAITREY MEHTA

Information Retrieval System, Selector System
maitrey@cs.utah.edu

VIVEK GUPTA

Question Typing, Information Extraction System
vgupta@cs.utah.edu

Thank you.

QUESTIONS?

Test it yourself:
https://bitbucket.org/vgupta123/simple-question-answering

References and Resources

[1] Cao J. , Song B. "Talking Geckos"

[2] Islam, A. and Inkpen, D., 2008. Semantic text similarity using corpus-based word similarity and string similarity. ACM Transactions on Knowledge Discovery from Data (TKDD), 2(2), p.10.

[3] Riloff, Ellen, and Michael Thelen. "A rule-based question answering system for reading comprehension tests." In Proceedings of the 2000 ANLP/NAACL Workshop on Reading comprehension tests as evaluation for computer-based language understanding systems-Volume 6, pp. 13-19. Association for Computational Linguistics, 2000.

Language of Implementation - Python 2.7.1
POS Tagger - NLTK
NER Tagger - SpaCy
CoreRef - Stanford CoreNLP
Dependency Parser - SpaCy
BIO Tagger - SpaCy