

Final Exam

601.467/667 Introduction to Human Language Technology

Fall 2020

Johns Hopkins University
Co-ordinator: Philipp Koehn

15 December 2020

Complete all questions.

Use additional paper if needed.

Time: 180 minutes.

Name of student: _____

Q1. Dialog Systems

20 points

1. Which (can be none or more than one) of the following are easy for today's conversational agents: [2 pts]
 - (a) question answering
 - (b) decision support
 - (c) meeting booking
 - (d) chitchat
2. Write down three ethical concerns / considerations regarding conversational agents (a.k.a. chatbots): [5 pts]
3. Describe 3 differences between task and non-task driven chatbot evaluation [3 pts]

4. True or False, if False explain why [10 pts]

(a) Machine translation is important for Chatbot Evaluation

True False

(b) Information Retrieval is important for chatbots.

True False

(c) Current trends show that chatbots are in decline.

True False

(d) Chitchat is useless and shouldn't be studied.

True False

(e) Companies are all using neural methods now.

True False

(f) Entity linking is important for persona.

True False

(g) Mutual information is important for neural generative chatbot models since this method takes into account the probability of the source given the target and the probability of the target given the source.

True False

(h) The mutual information objective decreases the likelihood of so-called dull responses.

True False

(i) Chatbots are attempting to include world knowledge, personal knowledge, and persona. This can only be done with transformer models.

True False

(j) Rule-based chatbots are a good choice when starting a project.

True False

Q2. Question Answering

20 points

1. Below is a canonical system diagram for factoid question answering. [12pts]

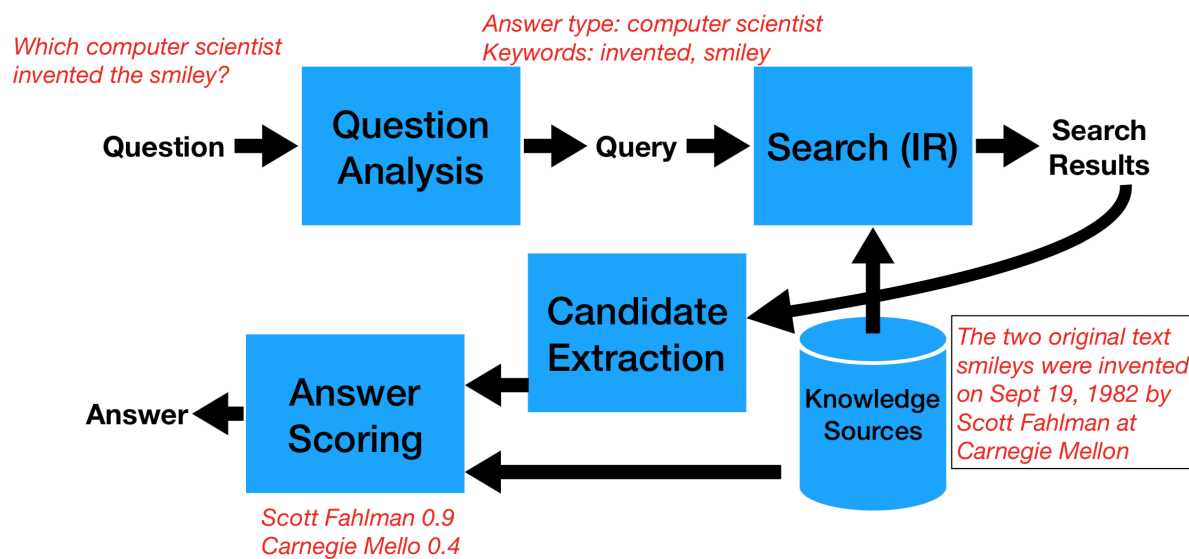
Describe in a sentence each what is the purpose of each component:

(a) Question Analysis:

(b) Search (IR):

(c) Candidate Extraction:

(d) Answer Scoring:



2. Explain the similarities and differences in the problem definition of

- Question Answering
- Machine Reading Comprehension
- Information Retrieval

3-4 sentences in total are sufficient. [8pts]

Q4. Clinical NLP

20 points

1. What is secondary use of medical records? [8 pts]

2. Assume you have a dataset of clinical notes (text). You remove all metadata that contains personally identifiable information. Can you publicly release the resulting dataset? Why not? [8 pts]

3. Named entity recognition (NER) is an NLP task common across many domains of data, but the types of entities often differ. Which of these entities would a clinical NER system identify? [4 pts]
 - Medications
 - Organizations
 - Locations
 - Citation
 - Treatments
 - Geopolitical entities
 - Symptoms
 - Author names

Q5. Ethics

20 points

1. Institutional Review Boards and Informed Consent:

(a) Describe what Institutional Review Boards are. [5 pts]

(b) What is the purpose of the Informed Consent in human subject research? [5 pts]

2. Ethical Principles

(a) Why do we need to ensure that AI algorithms take into account the principle of justice? How could considering fairness in AI affect the algorithm's output? [5 pts]

(b) Explain the concept of 'Explicability' in AI ethical principles and how it is related to the other four principles. (This question is referred to the five principles we saw in class, as explained in: Floridi L, Cowls J (2019): *A unified framework of five principles for AI in society*. Harvard Data Science Review.) [5 pts]