QA

Natural Language Processing

University of Maryland

Exercises

RAG Evaluation

You're brought in as a consultant to help build up a new question answering system for new mothers from diverse socio-economic backgrounds. The company is going to use machine reading to answer the questions.

RAG Evaluation

You're brought in as a consultant to help build up a new question answering system for new mothers from diverse socio-economic backgrounds. The company is going to use machine reading to answer the questions.

What source documents would you use? Where would they come from? How would you ensure that they questions retrieve the appropriate documents?

RAG Evaluation

You're brought in as a consultant to help build up a new question answering system for new mothers from diverse socio-economic backgrounds. The company is going to use machine reading to answer the questions.

You've now deployed the system, how do you ethically and safely evaluate whether it is working correctly? What metrics would you prioritize? If you use human annotators, describe your training and recruiting process. If you use an existing dataset, describe how you make sure it's appropriate for this task.

Turing Test

The next question is about the Turing Test. As you know, an interrogator has a limited ammount of time to determine whether their communication partner is a human or a computer. Their communications will be through writing questions at a terminal for ten minutes for each of the conversation partners. After each question, the conversation partner has 30 seconds to reply (to make up for typing speed). You know nothing about the potential human conversation partner other than that they're a "normal" person with a fair amount of education.

Turing Test

The next question is about the Turing Test. As you know, an interrogator has a limited ammount of time to determine whether their communication partner is a human or a computer. Their communications will be through writing questions at a terminal for ten minutes for each of the conversation partners. After each question, the conversation partner has 30 seconds to reply (to make up for typing speed). You know nothing about the potential human conversation partner other than that they're a "normal" person with a fair amount of education.

You're in charge of prepping advice for an interrogator who is clueless about computers and artificial intelligence. You have this page to offer advice: what would you tell them so that they wouldn't be fooled by the computer?

Turing Test

The next question is about the Turing Test. As you know, an interrogator has a limited ammount of time to determine whether their communication partner is a human or a computer. Their communications will be through writing questions at a terminal for ten minutes for each of the conversation partners. After each question, the conversation partner has 30 seconds to reply (to make up for typing speed). You know nothing about the potential human conversation partner other than that they're a "normal" person with a fair amount of education.

Let's now flip the script. You are developing an AI that will go up against a not-sosavvy interrogator. What strategies would you use to hold up as long as possible against their questions?

Cranfield vs. Manchester

For each of the following scenarios, identify how and why it can be both Manchester and Cranfield and give an example of a question or circumstance (e.g., the same question could be Manchester or Cranfield depending on what the question asker knows) from each paradigm given the scenario.

Cranfield vs. Manchester

For each of the following scenarios, identify how and why it can be both Manchester and Cranfield and give an example of a question or circumstance (e.g., the same question could be Manchester or Cranfield depending on what the question asker knows) from each paradigm given the scenario.

- A moderator at a political debate
- Police interrogating a suspect
- An consular officer processing a naturalization application
- A child asking their parents about the tooth fairy
- A customer asking a waiter about the special

Hard Question Examples

This question requires finding questions from the quiz bowl archive (`https://www.quizbowlpackets.com/`) which is designed to be pyramidal for humans. (You can also pull examples from the dataset you've been using for the homework.)

Hard Question Examples

This question requires finding questions from the quiz bowl archive (`https://www.quizbowlpackets.com/`) which is designed to be pyramidal for humans. (You can also pull examples from the dataset you've been using for the homework.)

Find a question that is pyramidal for a (typical) human but is not for a retrieval-based machine reading system. Rank the clues by difficulty for both humans and machines and explain when and why pyramidality is violated.

Hard Question Examples

This question requires finding questions from the quiz bowl archive (`https://www.quizbowlpackets.com/`) which is designed to be pyramidal for humans. (You can also pull examples from the dataset you've been using for the homework.)

Find a question clue that requires multihop reasoning. What are the evidence pieces that are required to solve the question?

Hard Question Examples

This question requires finding questions from the quiz bowl archive (`https://www.quizbowlpackets.com/`) which is designed to be pyramidal for humans. (You can also pull examples from the dataset you've been using for the homework.)

Find a question that would be very difficult to answer with a span extraction approach and explain why the approach would fail.