NATURAL LANGUAGE PROCESSING

LESSON 10 : QUESTION ANSWERING AND DIALOGUE SYSTEMS

*This lecture note contains excerpts from the following book.
Daniel Jurafsky & James H. Martin, Speech and Language Processing, 2019.

OUTLINE

- **Question Answering**
  - Overview of Question Answering
  - Types of Question Answering
  - Question Answering Components
  - Question Answering in TREC
- **Dialogue Systems**
  - Overview of Dialogue Systems
  - Types of Dialogue Systems
  - Dialogue System Components
  - Example Systems

OVERVIEW OF QUESTION ANSWERING

- Question answering (QA) is a computer science discipline within the fields of information retrieval and natural language processing, which is concerned with building systems that automatically answer questions posed by humans in a natural language.
- QA implementation, usually a computer program, may construct its answers by querying a structured database of knowledge or information, usually a knowledge base. More commonly, question answering systems can pull answers from an unstructured collection of natural language documents.

OVERVIEW OF QUESTION ANSWERING

QA studies attempt to address a wide variety of question types. These can be divided into themes such as fact, list, definition, how, why, hypothetical, semantically constrained, and cross-linguistic questions. However, we can divide QA systems into two according to the target data source.

- Closed-domain QA
- Open-domain QA

OVERVIEW OF QUESTION ANSWERING

In closed-domain Question Answering,

- The goal is to retrieve answers to questions within a specific domain.
- The main challenge is to develop a model that only requires small datasets for training.
- They might refer to a situation where only a limited type of questions are accepted, such as questions asking for descriptive rather than procedural information.

OVERVIEW OF QUESTION ANSWERING

In open-domain Question Answering,

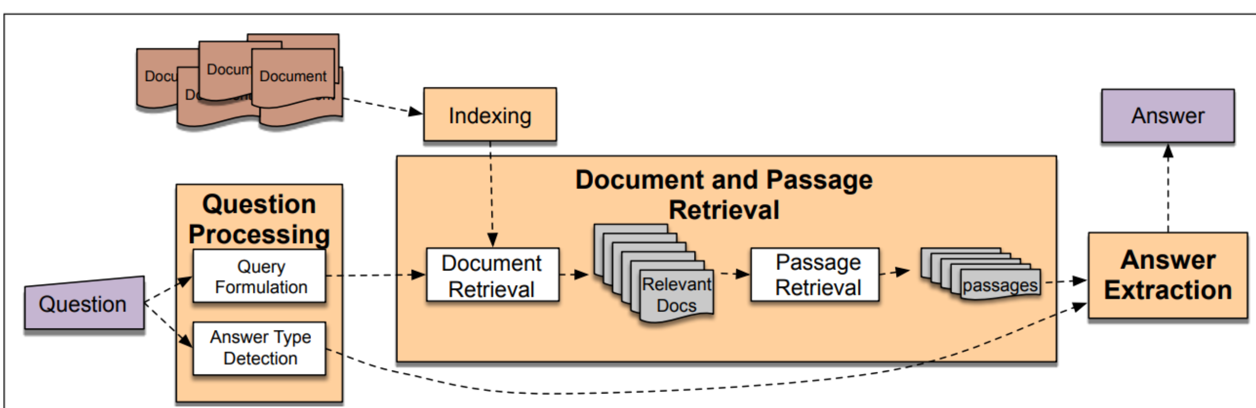
- The goal is to ask a model to produce answers to factoid questions in natural language.
- It deals with questions about almost anything and can use only general world knowledge that is not specific to a field.
- The true answer is objective, so it is simple to evaluate model performance.
- The real challenge is managing big data.

OVERVIEW OF QUESTION ANSWERING

By the early 1960s, systems used the two major paradigms of question answering:

- **IR based question answering**
- **Knowledge based question answering**

IR-BASED QUESTION ANSWERING



QUESTION PROCESSING STEP

The goal of the question-processing step is to extract the query, well the keywords passed to the IR system to match potential documents.

This leads to the following two steps:
- Classification of Question or finding answer type
- Query formulation or obtaining a list of keywords and tokens from the question

QUESTION PROCESSING STEP

Which US state capital has the largest population?

↓

The query processing might produce results like the following:

query: “US state capital has the largest population”

answer type: city

focus: state capital

QUERY FORMULATION STEP

Query formulation is the task of creating from the question a list of keywords that form a query that can be sent to an information retrieval system.

<Wh-word> did <A> verb <B> → ... <A> verb+ed <B>

Where is Ankara? → Ankara is located in Turkiye

When was the laser invented? → The laser was invented in 1960

QUESTION CLASSIFICATION STEP

The task of **question classification** is to determine the answer type.

- If the question starts with Who or Whom, QType is **PERSON**.
- If the question starts with Where, QType is **LOCATION**.

In general, question classification accuracies are relatively high on easy question types like PERSON, LOCATION, and TIME questions; detecting REASON and DESCRIPTION questions can be much harder.

DOCUMENT AND PASSAGE RETRIEVAL STEP

To rank passages by using supervised learning can be used following features:

- The number of **named entities** of the right type in the passage
- The number of **question keywords** in the passage
- The longest exact sequence of question keywords that occurs in the passage
- The rank of the document from which the passage was extracted
- The **proximity** of the keywords from the original query to each other
- The **N-gram overlap** between the passage and the question

ANSWER EXTRACTION STEP

The final stage of question answering is to extract a specific answer from the passage.

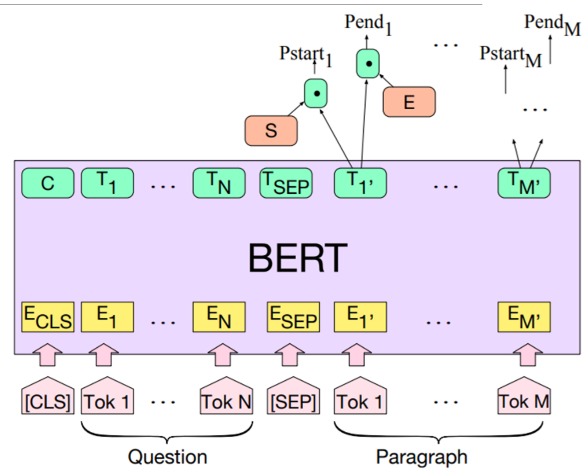
"How tall is Mt. Everest?"

The official height of Mount Everest is 29,029 feet

Unfortunately, the answers to many questions, such as DEFINITION questions, don't tend to be of a particular named entity type. For this reason modern work on answer extraction uses more sophisticated algorithms, generally based on supervised learning.

BERT BASED QUESTION ANSWERING

The power of contextual embeddings allow question answering models based on BERT contextual embeddings and the transformer architecture to achieve even higher accuracy.



KNOWLEDGE-BASED QUESTION ANSWERING

- We use the term knowledge-based QA for the idea of answering a question by mapping it to a query over a structured database.
- Systems for mapping from a text string to any logical form are called semantic parsers.
- Semantic parsers for QA usually map either to some version of predicate calculus or a query language like SQL.
- One of the first KB-QAs is BASEBALL about baseball game in 1961.

KNOWLEDGE-BASED QUESTION ANSWERING

The simplest formation of the knowledge-based question-answering task is to answer questions asking one of the missing arguments in a triple. Consider the example as follows:

Subject	Predicate	object
Ada Lovelace	birth-year	1815

Question	Logical Form
When was Ada Lovelace born?”	birth-year (Ada Lovelace, ?x)

EVALUATION OF QAs

- A common evaluation metric is Mean Reciprocal Rank (MRR) introduced in the Text Retrieval Conferences (TREC).
- MRR assumes a test set of questions that have been human-labeled with correct answers.
- MRR also assumes that systems are returning a short ranked list of answers or passages containing answers.
- Each question is then scored according to the reciprocal of the rank of the first correct answer.

EVALUATION OF QAs

- For example if the system returned five answers but the first three are wrong and hence the highest-ranked correct answer is ranked fourth, the rank score for that question would be 1/4.
- Questions with return sets that do not contain any correct answers are assigned a zero.
- A system's score, calculated by **MRR**, is the average of the scores it has earned for each question in the set.

$$\text{MRR} = \frac{1}{N} \sum_{i=1 \text{ s.t. } rank_i \neq 0}^{N} \frac{1}{rank_i}$$

OVERVIEW OF DIALOGUE SYSTEMS

- The main difference between QA and Dialogue Systems is the way they process and respond to user input.
- QA systems are designed to provide specific, concise answers to direct questions, often based on a pre-defined set of facts or data.
- Dialogue Systems, on the other hand, are typically more conversational in nature and are able to carry on a back-and-forth conversation with users.

TYPES OF DIALOGUE SYSTEMS

These programs generally fall into two classes.

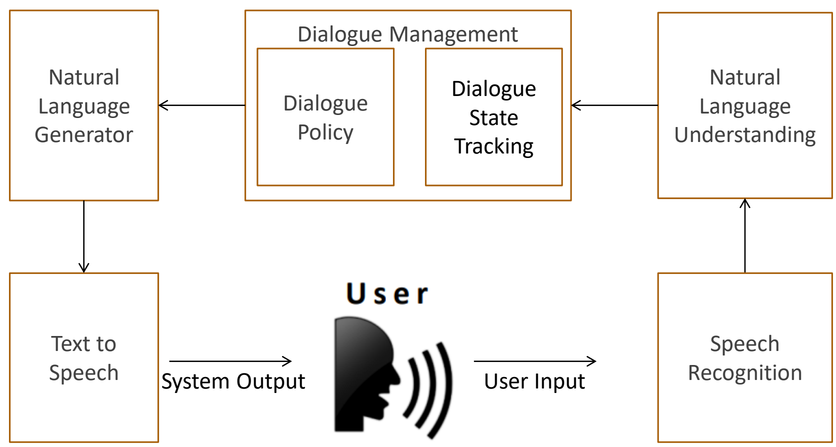
- **Task-oriented dialogue agents:** are designed for a particular task and set up to have short conversations .
- **Chatbots:** are systems designed for extended conversations.

TYPES OF DIALOGUE SYSTEMS

Dialog systems can also be classified as text-based and spoken-based systems.

- **In text-based systems,**
  Input and output are all text
- **In spoken-based systems,**
  there is a system that can speak with a voice and listen your voice

DIALOGUE SYSTEM COMPONENTS



NATURAL LANGUAGE UNDERSTANDING

The umbrella term "natural-language understanding" can be applied to a diverse set of computer applications, ranging from small, relatively simple tasks such as short commands issued to robots, to highly complex endeavors such as the full comprehension of newspaper articles or poetry passages.

Many real-world applications fall between the two extremes, for instance text classification for the automatic analysis of emails and their routing to a suitable department in a corporation does not require an in-depth understanding of the text, but needs to deal with a much larger vocabulary and more diverse syntax.

NATURAL LANGUAGE UNDERSTANDING

One simple NLU consists of:

- **dialogue act type** - encodes the user intention
- **filling slots** - describes entities in dialogue

«Is there maybe a cheap place in the center of town please?»

inform (price = cheap, area = center)

dialogue act type → inform

slots and values → price = cheap, area = center

DIALOGUE STATE TRACKING

- Dialogue state tracking is the process of keeping track of the information exchanged between two or more participants in a conversation.
- This can include the topic of the conversation, the goals and intentions of the participants, and any relevant background information.
- Dialogue state tracking is a key component of natural language processing, as it allows conversational agents to understand the context and content of a conversation and respond appropriately.

DIALOGUE STATE TRACKING

User: I'm looking for a cheaper restaurant
`inform(price=cheap)`
System: Sure. What kind - and where?
User: Thai food, somewhere downtown
`inform(price=cheap, food=Thai, area=centre)`
System: The House serves cheap Thai food
User: Where is it?
`inform(price=cheap, food=Thai, area=centre); request(address)`
System: The House is at 106 Regent Street

DIALOGUE POLICY

- The system needs to decide what dialogue act (type + slots) to realize in its next utterance.
- The next act may depend on the current dialogue state, the dialogue history, the confidence the system has, or the cost of making a mistake.
- Modern dialogue systems often make mistakes. It is therefore important for dialogue systems to make sure that they have achieved the correct interpretation of the user's input. This is generally done by two methods: confirming understandings with the user and rejecting utterances that the system is likely to have misunderstood.

NATURAL LANGUAGE GENERATION

- The NLG constructs the message that is to be spoken to the user.
- The NLG component takes the semantic representation of a communicative act from the system and generates a textual representation,
- Requires the making of decision regarding: Content planning (what to say) and Sentence realization (how to say it)

Input	Output
inform(phone=x)	The phone number is x

EXAMPLE SYSTEMS - Eliza

- Written by Joseph Weizenbaum in the mid-1960s as a simulation of a psychotherapist.
- User types questions, and ELIZA responds.
- Produce response based on analysis of input

- Transforms input to change the point of view (instead I you)

- Performs pattern recognition

- Substitutes keywords into canned phrases, or submits a general canned phrase

EXAMPLE SYSTEMS - Eliza

User: Men are all like
ELIZA: In what way?
User: They're always bugging us about something
ELIZA: Can you think of a specific example?
User: Well, my boyfirend made me come here.
ELIZA: Your boyfriend made you come here.
User: Yes, he says I'm depressed much of the time.

EXAMPLE SYSTEMS - AutoTutor

- It poses questions or problems that require approximately a paragraph of information to answer.
- Latent semantic analysis
- It has Curriculum script repository: ideal answer, misconceptions
- Corpus of documents: textbook
- Glossary: technical terms and definitions
- Dialogue management: finite state machine

EXAMPLE SYSTEMS - ChatGPT

- It uses Transformer-based GPT-3.5 technology.
- You can ask almost anything you can find on search engines.
- You can ask him to write code or even ask him to find bugs in your own codes.
- You can ask him to prepare an article, story or dissertation with some details.
- Many language support is available, including Turkish and English.