

# Knowledge-Based Question Answering

Sudipto Ghosh

*M.Sc. CS Semester I  
Department of Computer Science  
University of Delhi*

December 25, 2022

# Overview

## 1 Problem Statement

- Problem Definition
- Motivation

## 2 Modules

- KB Construction
- Question Understanding
- Inferencing Engine

## 3 System Overview

- Use Case Diagram
- System Architecture

# Problem Definition

- Parse unstructured text from domain corpus, identify entities, extract relations, map relations to domain concepts and build knowledge base.  
**Input:** Corpus, domain ontology, and training examples consisting of entity boundaries, relationship dependencies, valid triples.  
**Output:**  $\langle s, p, o \rangle$  triples to populate the knowledge graph.
- Model natural language question into a query, infer the facts about it required for the answer, assemble the facts into a natural language answer, and present it to the user.  
**Input:** Question as a spoken utterance or text prompt.  
**Output:** Answer/fact as a spoken utterance or text response.
- Ensure system maximizes performance on giving correct answers to a set of competency questions.

# Motivation

- Leveraging domain knowledge to improve virtual assistants.
- Inferencing step  $\implies$  system can answer unanticipated questions.
- Level of detail can be controlled to suit the expertise of the user.
- Can we answer complex questions that contain multiple subjects, express compound relations, or require simulated thinking?

## Example

**Q:** What is the capital of India?

**A:** The capital of India is New Delhi.

**Q:** What is the state of motor 2?

**A:** Motor 2 is currently turned off.

# KB Construction

- Knowledge base construction (KBC) is the process of populating a knowledge base (KB) with facts (or assertions) extracted from data.

## Sentence

Paracetamol, also known as acetaminophen, is usually prescribed for treating fever

## Entity Recognition

Paracetamol, also known as acetaminophen, is usually prescribed for treating fever

## Relation Extraction

Paracetamol, also known as acetaminophen, is usually prescribed for treating fever

## Coreference Resolution

Paracetamol, also known as acetaminophen | Paracetamol is usually prescribed for treating fever

# KB Construction (contd.)

## Triple Extraction

<Paracetamol, known as, acetaminophen>

<Paracetamol, prescribed for, fever>

## Entity Linking

<<https://en.wikipedia.org/wiki/Paracetamol>, known as, <https://en.wikipedia.org/wiki/Acetaminophen>>

<<https://en.wikipedia.org/wiki/Paracetamol>, prescribed for, <https://en.wikipedia.org/wiki/Fever>>

## Ontology Mapping

prefix wikidata <https://www.wikidata.org/wiki/Property>

<<https://en.wikipedia.org/wiki/Paracetamol>, wikidata:P2561, <https://en.wikipedia.org/wiki/Acetaminophen>>

<<https://en.wikipedia.org/wiki/Paracetamol>, wikidata:P2176, <https://en.wikipedia.org/wiki/Fever>>

- The resultant triples are stored to a triple or RDF store like Blazegraph through semantic queries, or a graph database like Neo4J.

# Question Understanding

- Understanding natural language questions refers to the ability to break down a question into the requisite steps for computing its answer.
- Encode questions into low-dimensional vectors with contextual information?
- Calculate semantic similarity between questions and entities in KB?
- Detect and link entities in questions to those in KB and construct queries?

## Question

What drug is prescribed for treating fever?

## Parsed Query

<?, prescribed for, fever>

## Semantic Query

<?, <https://www.wikidata.org/wiki/Property:P2176>, <https://en.wikipedia.org/wiki/Fever>>

# Inferencing Engine

- KBQA Models learn Question Answering by using a QA corpus and a populated KB – uncertainty, incompleteness and noise are inevitable
- Probabilistic Inferencing  $\implies$  infer predicates from templates.
- Offline – learn the mapping between templates and predicates.
- Online – break questions down to simple questions, make use of probability distributions, calculate maximum likelihood.
- Entity distribution, template distribution, value (answer) distribution.
- Questions in actual interactions might be vague and unusual.
- Answer questions with entities/predicates matching the top confidence score.

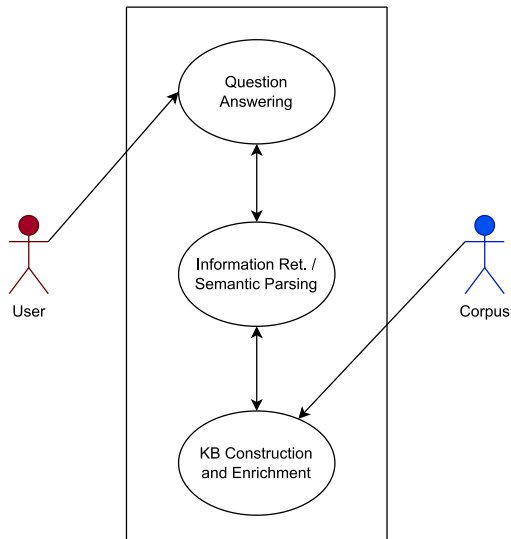
## Example

**Template:** what is **treated** by **\$medicine**?

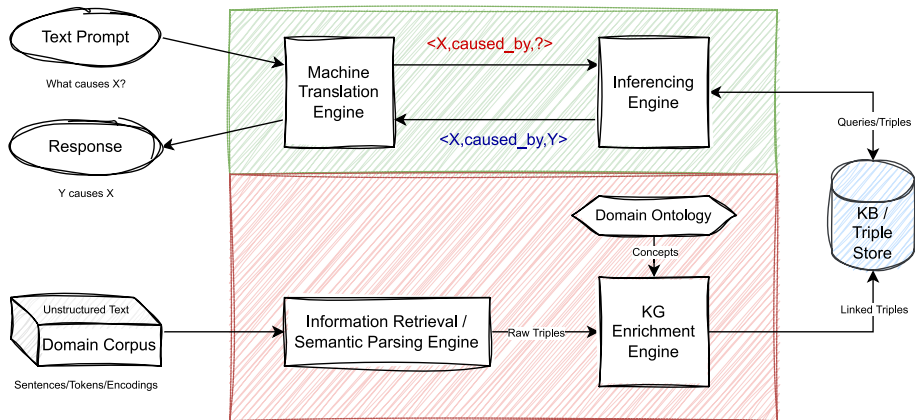
**Predicate:** **prescribed\_for** (maybe 90.5%), **founder** (maybe 0.01%)



# Use Case Diagram



# System Architecture



# References

- [1] Vasudha Bhatnagar et al. “Similar Cases Recommendation using Legal Knowledge Graphs”. In: *Proc. of KDD 2021*. DOI: 10.48550/ARXIV.2107.04771.
- [2] Sudipto Ghosh et al. “Constructing a Knowledge Graph from Indian Legal Domain Corpus”. In: *Proc. of TEXT2KG 2022 and MK 2022 co-located with ESWC 2022*, pp. 80–93. URL: [http://ceur-ws.org/Vol-3184/TEXT2KG\\_Paper\\_6.pdf](http://ceur-ws.org/Vol-3184/TEXT2KG_Paper_6.pdf).
- [3] Daniel Jurafsky and James H. Martin. “Question Answering”. In: *Speech and Language Processing*. USA: Prentice-Hall, Inc., 2021. Chap. 23. ISBN: 0131873210.
- [4] Yunshi Lan et al. “Complex Knowledge Base Question Answering: A Survey”. In: *Proc. of IJCAI 2021*. DOI: 10.48550/ARXIV.2108.06688.
- [5] Ce Zhang. “DeepDive: A Data Management System for Automatic Knowledge Base Construction”. PhD thesis. 2015. URL: <https://cs.stanford.edu/people/czhang/zhang.thesis.pdf>.

# Thank You