

## Question Answering

CISC489/689-010, Lecture #24

Wednesday, May 13<sup>th</sup>

Ben Carterette

## Question Answering

- Usual IR problem: submit query (usually keywords), receive ranking of documents
- QA: user asks a question in natural language, receives an answer in natural language or a ranking of answers
- Similar problems, but different in fundamental ways
  - Different approaches are successful
  - Different evaluation methods are needed



The screenshot shows the True Knowledge interface. At the top left is the logo "True Knowledge" with "LLC" in smaller text below it. To the right of the logo is a search input field containing the text "who is president". To the right of the input field is a button labeled "Answer". The rest of the interface is blank.

## Knowledge-Based Systems

- Create a database of known facts
- Reformulate question as a statement with a "blank"
  - Question: "who is president of the U.S.?"
  - Reformulation: "The president of the U.S. is \_\_\_\_"
- Find facts in database that match the statement
  - Fact in database: "The president of the U.S. is Barack Obama"
- If none match, find intermediate facts
  - For example, "The president of the U.S. is the former junior senator from Illinois", "Barack Obama is the former junior senator from Illinois" → "The president of the U.S. is Barack Obama"

## Reformulation

- Different reformulations may be possible depending on query type
  - “who” question: “\_\_\_ is the president of the U.S.”, “The president of the U.S. is \_\_\_”, “\_\_\_, who was born in \_\_\_, has been elected president of the U.S.”, ...
  - “when” question: “Barack Obama took office in \_\_\_”, “Barack Obama was president of the U.S. from \_\_\_”, ...
- Therefore it is useful to classify question by type
  - Who/what/where/when/how
  - Use type to generate reformulation patterns

## Question Classification

- What are some features we could use for this?
  - Does the question contain “who”/“what”/“where”/“when”?
  - Entity types: person names might imply “who”, place names might imply “where”, etc
- This is not always as easy as it might seem
  - “Name the first private citizen to fly in space” – what features could we use to determine that is a “who” question?