

CS474 Intro to Natural Language Processing

Question Answering

Question Answering



What was the name of the enchanter played by John Cleese in the movie "Monty Python and the Holy Grail"?



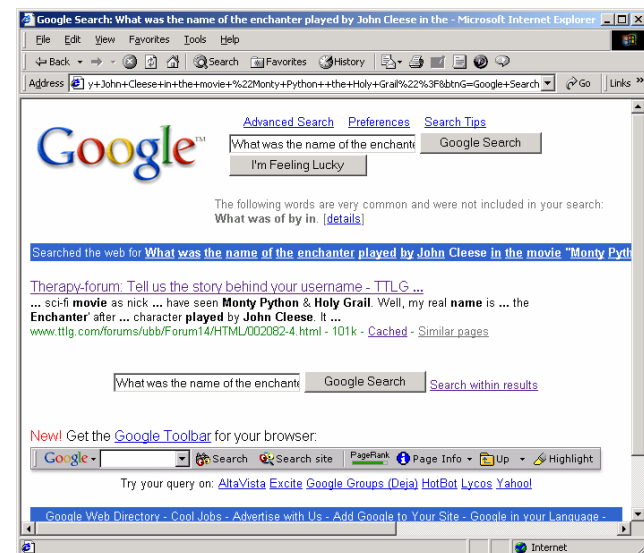
Trivia: John Cleese is a Cornell A.D. White professor-at-large.

Question answering

- Overview and task definition
- History
- Open-domain question answering
- Basic system architecture
- Predictive indexing methods
- Pattern-matching methods
- Advanced techniques

~2 lectures

IR solution



The document

The screenshot shows a forum thread on TTLG Forums. The browser window title is "Therapy-forum: Tell us the story behind your username - TTLG Forums - Microsoft Internet Explorer". The address bar shows a URL from google.com. The forum header includes navigation links like "Home", "Sshock2.com", "Thief-TheCircle.com", and "DeusEx-Machina.com". The thread title is "Therapy-forum: Tell us the story behind your username". The first post is by user "Nightvision" (Member, 540 posts, registered Mar 2000) with the subject "Darkblade: any relation of Mallus?". The second post is by user "Jacksberry" (Member, 226 posts, registered May 2000) with the text: "Helloe everybody, my name Jacksberry and i'm an escapist...i'm actually a fragile female thief and i've no idea why i named myself after a horny old man - but reading the scriptures and seeing the ghosts in TOB just made me so heartsick i".

The answer

The screenshot shows a reply to the forum thread. The browser window title is "Therapy-forum: Tell us the story behind your username - TTLG Forums - Microsoft Internet Explorer". The address bar shows a URL from google.com. The reply is from user "enchanter" (Member, 196 posts, registered Jan 2001) posted on February 13, 2001 at 03:09 AM. The text of the reply is: "Erm, well, I didn't come up with my own nick actually. I assume most of you have seen **Monty Python & Holy Grail**. Well, my real name is Tim and some of my strange friends that share my twisted sense of humour started calling me 'Tim the Enchanter' after the character played by John Cleese. It sort of stuck and got reduced to **enchanter**. I suppose it fits my personality well enough, although I know others use it on the 'net. By the way, not to rock the 'My fantasy author is better than yours' boat, but... N. How in the world can you say that Ro. Jordan is a bad writer? He's got to be one of the best fantasy writers since Tolkien, in my opinion. His characters are interesting (as you said) and believable and he is very good at interweaving plots. I've found his style very similar to Tolkien's in many ways, and I would be surprised to find anyone here who didn't say that Tolkien's *(The Hobbit)* books were fantastic".

Commercial QA systems

The screenshot shows the Ask Jeeves search engine interface. The browser window title is "Ask Jeeves". The search query is "What was the name of the enchanter played by John Cleese in Monty Python and the Holy Grail?". The search results are displayed below the query, showing several links related to the movie. A "HELPFUL HINT" section is also visible, advertising "Rent DVDs Online" with a free trial and free shipping. The search results list includes: "1. The DVD Journal: Monty Python and the Holy Grail, Special Edition".

Not so easy for people either...



Question answering

- Overview and task definition

➔ History

- Open-domain question answering
- Basic system architecture
- Predictive indexing methods
- Pattern-matching methods
- Advanced techniques

History

- **Closed-domain QA systems**
 - LUNAR [Woods & Kaplan, 1977]
 - WOLFIE [Thompson & Mooney, 1998]
 - Q/A [Lehnert, 1978]
- **Open-domain QA systems**
 - TREC QA evaluations [1999, 2000, 2001, ...]

LUNAR

- **Answered questions about moon rocks and soil gathered by the Apollo 11 mission**
 - Data base of information for all collected samples
- **Architecture**
 - Parse English question into a data base query
 - Syntactic analysis via augmented transition network parser and heuristics (including some for shallow semantics)
 - Semantic analysis maps parsed request into query language; query denotes unambiguous meaning of the request
 - Run query on data base to produce answer

LUNAR

- **Resources required**
 - Parser for a subset of English (size unclear)
 - Handled tense, modality, some anaphora, some conjunctions, some relative clauses, some adjective modifiers (dealing with quantification)
 - Vocabulary of about 3,500 words
- **Sample questions**
 - What is the average concentration of aluminum in high alkali rocks?
 - What samples contain P205?
 - Give me the modal analyses of P205 in those samples.


LUNAR example

- **Do any samples have greater than 13 percent aluminum?**

- **Data base query**

(TEST (FOR SOME X1 / (SEQ SAMPLES): class to test
T;
(CONTAIN
X1
(NPR* X2 / 'AL203)
(GREATER THAN 13 PCT))))

No restriction on class
Proposition



- **Answer:**

– Yes

LUNAR assessment

- **System characteristics**

- Closed domain (lunar geology and chemistry)
- Structured data (information contained in a data base)
- Structured answers (information contained in a data base)
 - Avoided dialogue problems
- Context: sophisticated users demanding high accuracy

- **Labor intensive to build**

- Complex system
- High accuracy required
- Few general-purpose NLP resources available at the time

LUNAR assessment

- **Research on systems like LUNAR continued for another decade**

- **Focused on**

- Syntactic parsing
- Incorporating domain knowledge
- Dialogue management

- **Problems**

- Expensive to build
- Brittle...prone to unexpected sudden failure

WOLFIE

- **Word Learning From Interpreted Examples**

[Thompson and Mooney, 1998]

- Closed domain, structured data, structured answers
- Avoids labor-intensive system development
- Uses *inductive logic programming* methods to acquire parsers that can map natural language queries into executable logical form (i.e. data base queries)
 - Requires examples of NL queries and their logical form
- Indirectly evaluate the parser based on the number of queries that system gets right/wrong.

Lehnert's Q/A system

- **Implemented a broader theory of question answering**
 - motivated by issues of cognitive plausibility
 - relied on the linguistic/cognitive theories of *conceptual dependency*, *scripts*, *plans*, etc. [Schank, 1970's]
 - used a question taxonomy
 - somewhat closed domain (actions), unstructured data, generated answers
 - answered questions about an arbitrary input text (usually event-based)

Q/A example 1

- **Input text**

John threw the baseball to Mary. She missed the ball and it hit her on the head.
- **Questions:**
 - Was Mary happy?
 - Who has the baseball?
 - Why did John throw the baseball to Mary?

Lehnert's Q/A system

- **“Parse” input text into a semantic representation (*conceptual dependency*)**
- **Generate inferences from that representation**
 - Inferences associated with CD's semantic primitives
- **“Parse” the question, mapping it into one of the predefined question types**
- **Employ the method associated with the question type to answer the question**

Q/A example 2

- **Input text**

For their first date, John took Mary to McDonald's for burgers. Mary was not impressed.
- **Questions:**
 - Did John and Mary pay for the burgers?
 - What did John and Mary eat for dinner on their first date?

Q/A assessment

- **Labor intensive to build**
 - Complex system
 - Background knowledge needed
 - Data structures to encode scripts, plans, goals, inferences associated with CD primitives
 - Few general-purpose resources available at the time
- **Not designed to be a general-purpose Q/A system**

Question answering

- **Overview and task definition**
- **History**
- ➔ **Open-domain question answering**
- **Basic system architecture**
- **Predictive indexing methods**
- **Pattern-matching methods**
- **Advanced techniques**

Towards open-domain QA

Which country has the largest part of the Amazon rain forest?

The chaotic development that is gobbling up the Amazon rain forest could finally be reined in with a new plan developed by officials of Amazon countries and leading scientists from around the world.

“That’s some of the most encouraging news about the Amazon rain forest in recent years,” said Thomas Lovejoy, a tropical ecologist at the Smithsonian Institution and an Amazon specialist.

“It contrasts markedly with a year ago, when there was nothing to read about conservation in the Amazon, especially in *Brazil*, except bad news,” Lovejoy said in a recent interview.

Sixty percent of the Amazon, the world’s largest tropical rain forest, *lies in Brazil*, but the forest also covers parts of the eight surrounding countries.

Lovejoy was one of the organizers of an unusual workshop held in mid-January in Manaus, *Brazil*, a sprawling city of 1 million people in the heart of the Amazon. It was the center of *Brazil*’s once-thriving rubber trade.

Question Answering

- **Simplifications**
 - short-answer, fact-based questions
 - answer exists in the collection as a text fragment
 - supporting info can be found in a single document
 - system returns up to 5 guesses per question
- **Sample questions**
 - How many calories are there in a Big Mac?
 - Who is the voice of Miss Piggy?
 - Who was the first American in space?
 - Where is the Taj Mahal?

TREC QA: evaluation

- **Human assessors judge the answers**
 - Allowed to accept multiple answers
- **Systems scored on *mean reciprocal rank of first correct answer***
 - First answer correct = 1 point, second answer correct = 1/2 point, third answer correct = 1/3 point, ...
 - 0 if none of the n answers are correct
 - Average across all questions
- **Also reported on the number of questions answered correctly**

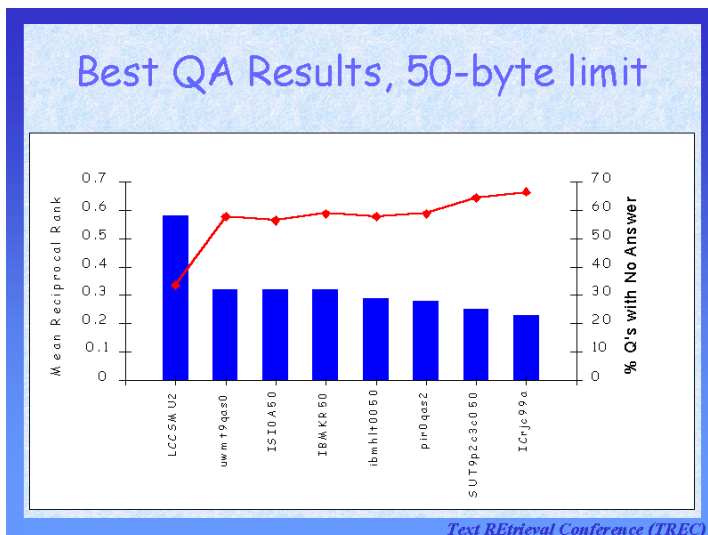
Question Answering

- **Performance**

	%correct	avg rank 1st
TREC-8 (1999):	70%	1.4
TREC-9 (2000):	65%	1.7
TREC-10 (2001):	70%	1.3
TREC-11 (2002):	83%	1
TREC-12 (2003):	70%	1
TREC-13 (2004):	84%	1

*harder
questions
1 guess, exact*

State of the art: TREC 2000



Courtesy of
E. Voorhees

List Questions

- **List questions**

1915: List the names of chewing gums.

Stimorol	Orbit	Winterfresh	Double Bubble
Dirol	Trident	Spearmint	Bazooka
Doublemint	Dentyne	Freedent	Hubba Bubba
Juicy Fruit	Big Red	Chiclets	Nicorette

- **Performance**

- TREC 2003: F .40
- TREC 2004: F .62

Definition Questions

- **Who is Colin Powell?**
- **What is mold?**

- **Audience?**

- **Evaluation requires matching *concepts* in the desired response to *concepts* in a system response**
 - TREC 2003: F .55

Context Task

- **Context with target item**

21	Club Med		
21.1	FACTOID	How many Club Med vacation spots are there worldwide?	
21.2	LIST	List the spots in the United States.	
21.3	FACTOID	Where is an adults-only Club Med?	
21.4	OTHER		

- **Performance**
 - TREC 2004
 - Factoids: .84 initial; .74 non-initial
 - Lists: .62 F
 - Other: .46 F

Question answering

- **Overview and task definition**
- **History**
- **Open-domain question answering**
- ➔ **Basic system architecture**
 - [Cardie et al., ANLP 2000]
- **Predictive indexing methods**
- **Pattern-matching methods**
- **Advanced techniques**