

Question answering

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

Question Answering

Question Answering

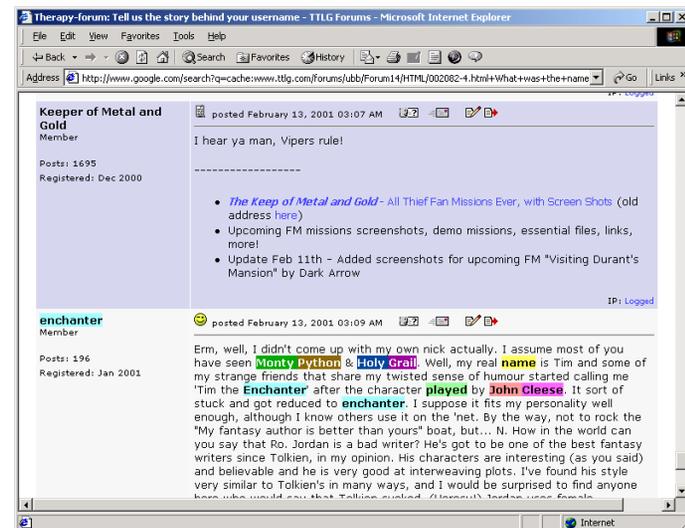


What was the name of the enchanter played by John Cleese in the movie “Monty Python and the Holy Grail”?



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

Answer from the Web (circa 2001)

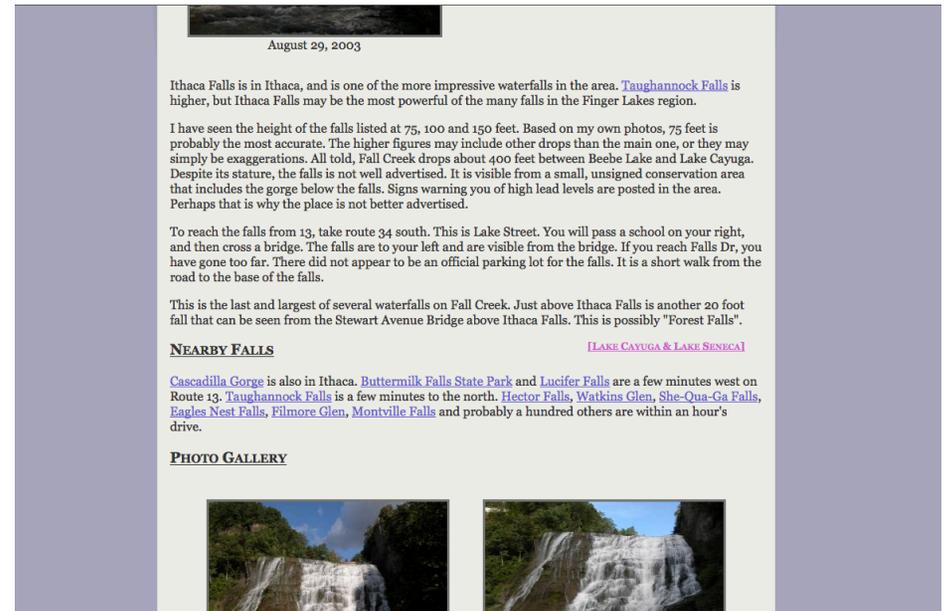
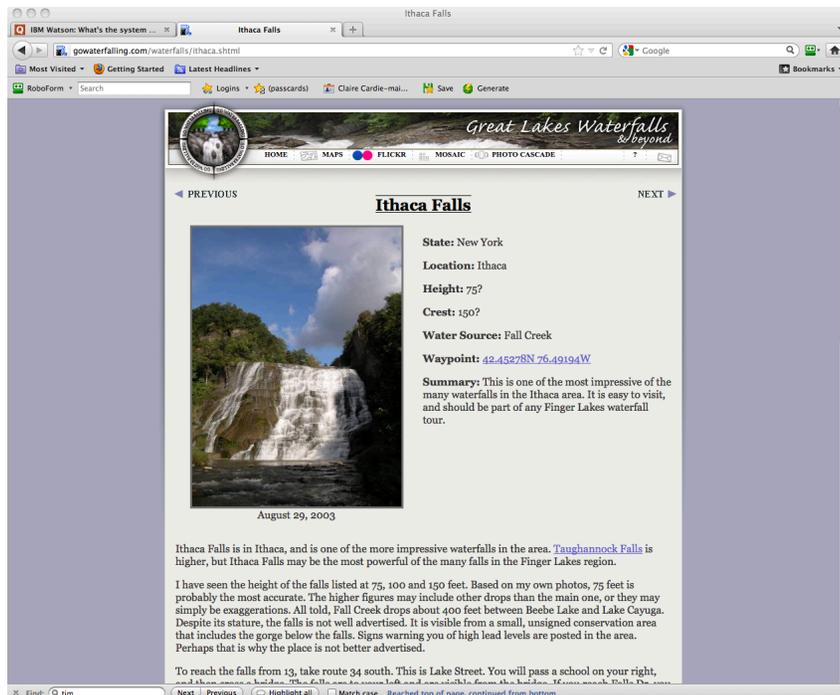
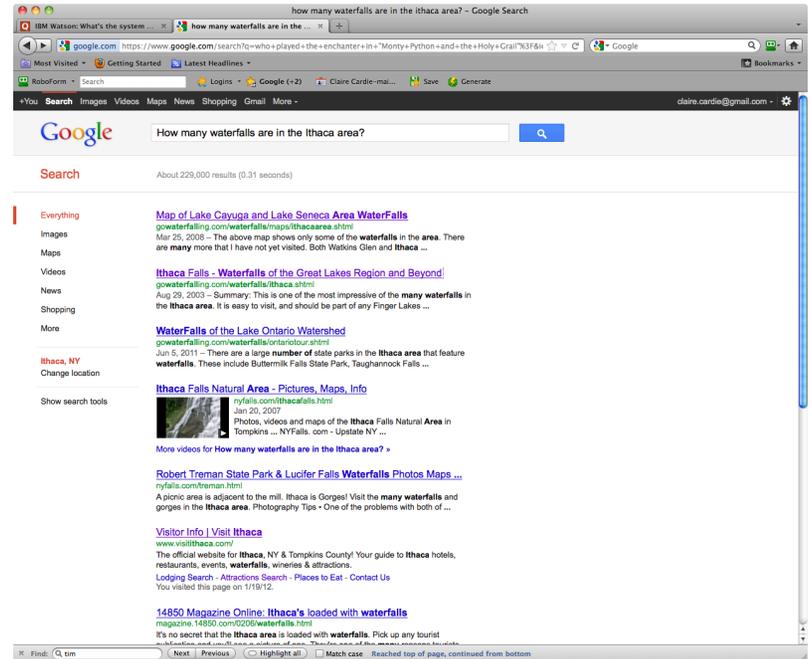


Top-ranked page TODAY

Tim the Enchanter is a [fictional character](#) from [Monty Python's 1975 movie](#) [Monty Python and the Holy Grail](#) played by [John Cleese](#). In the play [Spamalot](#) based on the movie, the part was originally played by [Hank Azaria](#). Tim is a strange reclusive [wizard](#) or [conjurer](#) who wears [ram horns](#) on his skullcap and has a penchant for [fire-based magic](#). Tim's overall presence is commanding, and observers of his magic often applaud after a display. He speaks with a [Scottish accent](#).

[video](#)

Wikipedia



Question answering

- Overview and task definition
- ➔ **History**
- **Open-domain question answering**
- **Basic system architecture**
 - Watson's architecture
- **Techniques**
 - 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



LUNAR

- **Answered questions about moon rocks and soil gathered by the Apollo 11 mission**
 - Data base of information for all collected samples
- **Architecture**
 - Transform 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.

Search: What is the average concentration of aluminum in high alkali rocks? (0.41 seconds)

Search results: About 2,550,000 results (0.41 seconds)

Everything: Advances in computers - Google Books Result

Images: books.google.com/books?isbn=0120121174...

Maps: Franz L. Alt, Marshall C. Yovits - 1978 - Computers - 333 pages

Videos: For example, the geologist should be able to ask questions such as "What is the average concentration of aluminum in high-alkali rocks?" without having to ...

News: Machinery of the mind: inside the new science of artificial ... - Google ...

Shopping: books.google.com/books?isbn=0307799395...

Books: George Johnson - 2011 - Computers

More: Geologists who wanted to ask something like, "What is the average concentration of aluminum in high alkali rocks?" would have to write a program that would ...

Ithaca, NY: The Handbook of Computational Linguistics and Natural Language ...

Change location: books.google.com/books?isbn=1405155817...

Show search tools: Alexander Clark, Chris Fox, Shalom Lappin - 2010 - Computers - 800 pages

IPPI 53 AI - Aluminium: Median concentrations of Al in the FOREGS samples and in some reference ... high A2O3 values are related to alkaline volcanic rocks. In stream sediment ...

IPPI Semantics and Quantification in Natural Language Question ...

The Geology of Rare Earth Elements

Google search (Jan 2012)

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 as much as possible
 - 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]



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

[Lehnert, 1978]

- **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 type 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**

Two distinct paradigms of research in NLP: CL and AI. (Re-united in the mid-1990' s.)

Question answering

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

TReC QA framework

- **Early simplifications (late 1990’s)**
 - 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?

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.

TReC QA: evaluation

- **Human assessors judge the answers**
 - Allowed to accept multiple answers
- **Systems scored on *mean reciprocal rank (MRR)* of first correct answer**
 - if first answer correct = 1 point,
 - else if second answer correct = ½ point,
 - else if third answer correct = 1/3 point, ...
 - 0 if none of the *n* answers are correct
 - Average of the *reciprocal rank* across all questions
- **Also reported on the number of questions answered correctly**

Question Answering

- **Performance**

	<u>%correct</u>	<u>avg rank 1st</u>	
TREC-8 (1999):	70%	1.4	
TREC-9 (2000):	65%	1.7	<i>harder questions</i>
TREC-10 (2001):	70%	1.3	<i>NIL answers</i>
TREC-11 (2002):	83%	1	<i>1 guess, exact</i>
TREC-12 (2003):	70%	1	
TREC-13 (2004):	84%	1	