Introduction to NLP

Question Answering
Siri

“What is the population of Sri Lanka?”

I found this for you:

Input interpretation
Sri Lanka population

Result
20.4 million people (world rank: 56th)
(2010 estimate)

Recent population history

Voice Control

Cancel
Who played Linus in Lost?

Ben Linus

Benjamin Linus is a fictional character portrayed by Michael Emerson on the ABC television series *Lost*. Ben was the leader of a group of island natives called the *Others* and was initially known as Henry Gale to the survivors. With Oceanic Flight 815. He began as the antagonist during the second and third seasons, but in subsequent seasons, becomes something of an uneasy ally to the main characters. Even so, however, throughout the series, he is persistently characterized by spontaneous coldblooded... Read More »

Go to: Ask Encyclopedie - Images - Videos


Related Questions:
- Spiritually speaking, who is better looking: Ben Linus, Henry Gale, or Michael Emerson?

More Q&A »

Michael Emerson - IMDb

Michael Emerson, Actor, Lost. Michael... He is an actor, known for *Lost* (2004), Person of Interest (2011) and *Saw* (2004). He has... Narrator / Benjamin Linus.

Roger Linus - Lostpedia - The Lost Encyclopedia - Wikia

Roger Linus was the father of Benjamin Linus, whom he regularly abused, both verbally and... Portrayed by... Outside references to Lost/Video games.
What is the area of a circle with radius 5 ft?

Input interpretation:
- circle radius 5 feet area

Result:
- $25 \times \pi^2 \text{ (square feet)} \approx 78.5398 \text{ft}^2 \text{ (square feet)}$

Visual representation:
- A circle with a radius of 5 feet.
IBM’s Watson

http://www.geekwire.com/2013/ibm-takes-watson-cloud/
Watson on Jeopardy

– Sample questions
  – On December 8, 2008 this national newspaper raised its newsstand price by 25 cents to $1
    • USA Today
  – In 2010 this former first lady published the memoir "Spoken From the Heart"
    • Laura Bush
  – This person is appointed by a testator to carry out the directions & requests in his will
    • Executor
  – Familiarity is said to breed this, from the Latin for "Despise"
    • Contempt
  – As of 2010, Croatia & Macedonia are candidates but this is the only former Yugoslav republic in the EU
    • Slovenia
  – The ancient "Lion of Nimrud" went missing from this city's national museum in 2003 (along with a lot of other stuff)
    • Baghdad
  – It's just a bloody nose! You don't have this hereditary disorder once endemic to European royalty
    • Haemophilia
  – It's Michelangelo's fresco on the wall of the Sistine Chapel, depicting the saved and the damned
    • The Last Judgement
  – She "Died in the church and was buried along with her name. Nobody came"
    • Eleanor Rigby
  – A camel is a horse designed by this
    • Committee
Watson Videos

• IBM's Watson Supercomputer Destroys Humans in Jeopardy | Engadget
  – https://www.youtube.com/watch?v=WFR3IOm_xhE
• "What is IBM Watson?" 7 Videos from the Jeopardy! Era
People Ask Questions Online

• Excite corpus
  – 2,477,283 queries (one day’s worth)

• 8.4% of them are questions
  – 43.9% factual (what is the country code for Belgium)
  – 56.1% procedural (how do I get out of debt) or other
The Excite Corpus

- In what year did baseball become an official sport?
- how do i get out of debt?
- Where can I found out how to pass a drug test?
- When is the Super Bowl?
- who is California's District State Senator?
- where can I buy extra nibs for a foutain pen?
- what time is it in west samoa?
- Where can I buy a little kitty cat?
- what are the symptoms of attention deficit disorder?
- Where can I get some information on Michael Jordan?
- How does the character Seyavash in Ferdowsi's Shahnameh exhibit characteristics of a hero?
- When did the Neanderthal man live?
- Which Frenchman declined the Nobel Prize for Literature for ideological reasons?
Murax Questions

- What U.S. city is at the junction of the Allegheny and Monongahela rivers?
- Who wrote “Across the River and into the Trees”?
- Who married actress Nancy Davis?
- What’s the capital of the Netherlands?
- Who was the last of the Apache warrior chiefs?
- What chief justice headed the commission that declared: “Lee Harvey Oswald . . . acted alone.”?
- What famed falls are split in two by Goat Island?
- What is November’s birthstone?
- Who’s won the most Oscars for costume design?
- What is the state flower of Alaska?
AOL Corpus Questions

- what does cerebral cortical atrophy mean
- what fraction is closest to pi
- what is the highest calories consumed by a person in a 24 hour period
- what happens to limewater in the presence of carbon dioxide
- what language is numa numa in
- who discovered Saturn
- who discovered the gene for Huntington's disease
- who is the daughter of Indian sitar maestro Ravi Shankar
- who invented marshmallow sandwich cookies
- who invented lithograph
- who was the first Michigan football player to put on the cover of a video game
- who invented smallpox vaccine
- when was the Berlin Wall torn down
- when was Yahoo launched
- when was the wedding march by Mendelssohn composed
- when was the Declaration of Independence signed
- when did France become a republic
- when was Alexander the Great born
- which fish contains low mercury
- when is the youngest a minor can move out of their parent's house in Texas
- when should a child see a psychologist
- what hormone imbalance could mimic pregnancy symptoms
Question Type Taxonomy

• Question Types
  – Yes/no vs. wh-
  – Factual vs. procedural
  – Single answer vs. multiple answers
  – Objective vs. subjective
  – Context-specific (personalized) vs. generic
  – Known answer in the collection, y/n
State of the Art

• Question types
  – Mostly factual, short-answer questions

• System architecture
  – IR component
  – Statistical approaches using lots of data
  – Relatively little knowledge
Older Systems (pre-TREC)

- **BASEBALL** (Green et al., 1961)  
  - Baseball
- **ELIZA** (Weizenbaum 1966)  
  - Dialogue, Therapist
- **SHRDLU** (Winograd 1972)  
  - Blocks world
- **LUNAR** (Woods, 1973)  
  - Lunar rocks
- **Murax** (Kupiec 1993)  
  - Open-domain, encyclopedia-based
- **START** (Katz 1997)  
  - Open-domain, web-based
- **Deep Red** (Hirschman & al. 1999)  
  - Reading comprehension
- **Jupiter** (Zue et al., 2000)  
  - Spoken questions
Eliza (Weizenbaum 1966)

• Eliza plays a therapist
  – and doesn’t really answer questions 😊
• Simple pattern matching
• Sample dialogue
  E: I am the psychotherapist. Please, describe your problems.
  P: I am always tired.
  E: Why do you say you are always tired?
  P: Because I cannot sleep.
  E: Is it because you cannot sleep that you came to me?
  P: Yes, help me please.
  E: Earlier you said you cannot sleep?
Lunar (Woods et al., 1972)

- Lunar Sciences Natural Language Information System
- Apollo 11 moon rocks
- Questions in English
- Components
  - a general-purpose ATN (augmented transition network) grammar for English
  - a procedural-semantics framework for semantic interpretation
  - mapping from the syntactic grammar to a logic representation
  - 3,500 word dictionary
  - 13,000 entries in the knowledge base
- Lunar answered 78% of sentences asked by geologists at 1971 Lunar Rocks conference
  - How many breccias contain olivine?
  - What are they?
Lunar Grammar Fragment
**SHRDLU (Winograd 1972)**

- Simple questions about the blocks world
  - Person: Pick up a big red block.
  - Computer: OK.
  - Person: Grasp the pyramid.
  - Computer: I don't understand which pyramid you mean.
  - Person (changing their mind): Find a block which is taller than the one you are holding and put it into the box.
  - Computer: By "it", I assume you mean the block which is taller than the one I am holding.
  - Computer: OK.
  - Person: What does the box contain?
  - Computer: The blue pyramid and the blue block.
  - Person: What is the pyramid supported by?
  - Computer: The box.
  - Person: How many blocks are not in the box?
  - Computer: Four of them.
  - Person: Is at least one of them narrower than the one which I told you to pick up?
  - Computer: Yes, the red cube.

http://hci.stanford.edu/~winograd/shrdlu/
Start (Katz 1997)

- Open-domain
- Uses the Web
- Online since 1993

http://start.csail.mit.edu