

Question Answering

NLP Seminar, Fall 2009, TAU

A presentation by Ariel Stolerman Instructor: Prof. Nachum Dershowitz

Overview

- What is Question Answering?
- Approaching QA
 - Answer extraction
- A brief history
- Main Domains
 - NL database systems
 - Dialog systems
 - Reading comprehension systems
 - Open domain QA
- TREC
- Question answering based on semantic graphs



What is Question Answering?

Main aim:

present the user with a short answer to a question rather than a list of possibly relevant documents.

▶ Difficult finding answers on the WWW → QA Tech becomes more relevant

But what is a good answer?

Approaching Question Answering

Domains:

- Closed domain QA: dealing with questions under a specific domain
- Open domain QA: dealing with questions about basically everything



General approach:

- Analyze and classify the question
- Using information retrieval techniques, construct a subset of available relevant documents
- Analyze the retrieved documents, search for wanted entities
- If found, return the entity as a response, i.e. extract answer

Where do answers come from???

- Question reformulation pattern
- Word matching
- Question Classification
- Semantic parsing:
 - Type matching (role labeling)
 - Relation matching
 - Logical inference
 - Ranking pairs inference (hypernyms, synsets)
 - WSD
 - World knowledge
 - Coreference resolution
 - **...**
- Definitions
- Clarifications from the user

Q: What is a dog?

A: A dog is an animal...

Relational qtargets - R-CAPITAL-PLACE - What is California's

capital?

Abstract qtargets - A-DEFINITION,

A-WHY-FAMOUS-PERSON - Who

was Jane Goodall?

Was WalMart sued for sexual discrimination?

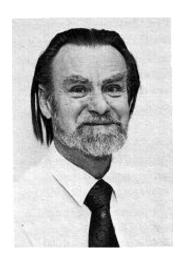
T: WalMart defended itself in court today against claims that its female employees were kept out of jobs in management because they are women

H: WalMart was sued for sexual

discrimination

A brief history of QA

- Not a new field of research Simmons, 1965
 "have a conversation with a book"
- Can be found in many areas of NLP research:
 - Natural language database systems
 - Dialog systems
 - Reading comprehension systems
 - QA in open domains

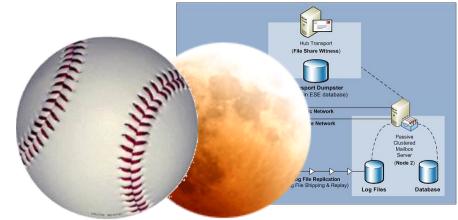


Natural Language Database Systems

Analyzing questions to produce a database query "list all players who scored more than 600 points in all games in 2009"

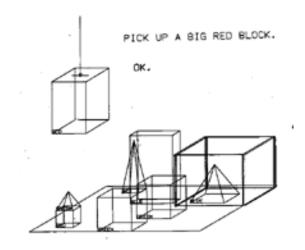
```
select player_name
from players
where (total score > 600) and (year=2009)
```

- From the first QA systems to be created
- Examples:
 - BASEBALL
 - LUNAR



Dialog Systems

- Conversation with a computer
 - Answering questions
- Examples:
 - SHRDLU
 - ▶ ELIZA
 - chatterbots



Person: PICK UP A BIG RED BLOCK.

Computer: OK. (does it)

Person: GRASP THE PYRAMID.

Computer: I DON'T UNDERSTAND WHICH PYRAMID YOU MEAN.
Person: 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. (does it)

Person: WHAT DOES THE BOX CONTAIN?

Computer: THE BLUE PYRAMID AND THE BLUE BLOCK.

Reading Comprehension Systems

- Used to test reading level of school children
- Using reading comprehension systems to test language understanding of computer systems
- Examples:
 - QUALM
 - Quarc
 - Deep Read



How Maple Syrup is Made

Maple syrup comes from sugar maple trees. At one time, maple syrup was used to make sugar. This is why the tree is called a "sugar" maple tree. Sugar maple trees make sap. Farmers collect the sap. The best time to collect sap is in February and March. The nights must be cold and the days warm. The farmer drills a few small holes in each tree. He puts a spout in each hole. Then he hangs a bucket on the end of each spout. The bucket has a cover to keep rain and snow out. The sap drips into the bucket. About 10 gallons of sap come from each hole.

- Who collects maple sap? (Farmers)
- What does the farmer hang from a spout? (A bucket)
- When is sap collected? (February and March)
- Where does the maple sap come from? (Sugar maple trees)
- Why is the bucket covered? (to keep rain and snow out)

Open Domain QA

- No specific domain restrictions
- Use of large text collections
- The WWW is the most popular text collection source for open domain systems
- Pros and Cons:
 - No limitation = lots of information / false information
 - Dated information / Outdated information



TREC - Text Retrieval Conferences

- Encouraging IR research areas since 1992
- Numerous tracks exploring numerous research fields:
 - Genomic track
 - Legal track
 - **...**
 - Question Answering
- Increasing difficulty from year to year

TREC QA track: http://trec.nist.gov/data/qamain.html



from large text collections.



Question Answering based on Semantic Graphs

Lorand Dali, Delia Rusu, Blaž Fortuna, Dunja Mladenić & Marko Grobelnik

> Department of Knowledge Technologies, Jožef Stefan Institute, Slovenia

Presented at the WWW conference, Madrid, April 2009

Motivation

FEATURE - Chinese medicine threatens Sumatran tiger.

Two of Indonesia's three species of tigers have become extinct and the third is severely threated because of the belief that tiger parts can improve sexual performance, experts say.

According to traditional Chinese medical belief, mental problems, laziness, malaria, epilepsy, toothache and lack of sexual appetite can be treated with tiger parts, leading to rampant p "Poaching is changing all the time...this trend for tiger products is new. There are people wanting tiger products who didn't want them before," Ron Lilley, coordinator for species conservable said tigers in Indonesia, now reduced to about 500 animals on the island of Sumatra, are particularly vulnerable. Poaching is a severe threat, with huge markets for tiger bones in "This is...perhaps a billion-dollar industry, mostly illegal and run by people not interested in sustainability," he said.

According to South Korean immigration statistics, the country imported 3,994 kilograms (8787 pounds) of tiger bones from Indonesia between 1970 and 1993, making Indonesia their That's a lot of tigers, Lilley said, since the bones of one tiger weigh less than 10 kilograms (22 pounds).

Speaking this week after the announcement of a joint venture with photo equipment manufacturer Kodak, which has pledged 50 million rupiah (\$21,000) to saving the Sumatran tiger, L "Our estimate is that 14 tigers are killed every year...maybe more. A population of less than 100 tigers has a very low possibility of surviving for the next 50 years because there can be n Extinction of the tiger is a real fear in Indonesia, which has seen the Bali tiger die out in the 1940s and the Java tiger disappear in the late 1980s.

Despite concerns over poaching, the biggest threat to the Sumatran tiger is the destruction of its habitat, Lilley said.

"Tiger populations have crashed without exception in all areas of the world," Lilley said. In Indonesia, this was "in the balance because conservation areas are being eaten into."

The WWF estimates the world tiger population at about 7,200. There are around 3,750 Bengal tigers left in India and just under 2,000 tigers in Southeast Asia, excluding Indonesia.

Subijanto, a spokesman for the Forestry Ministry, said Indonesia was committed to protecting the tigers, which live within Sumatra's four designated conservation areas.

The Kerinci conservation project, an area of around three million hectares (7.4 million acres) in west Sumatra, was being supported by funds from the World Bank, Subijanto said.

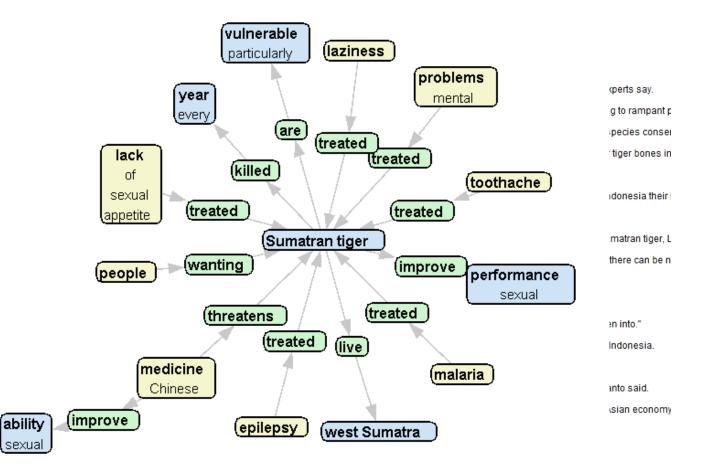
Whilse the Sumatran tiger is a fully protected species under the 1990 Conservation Act, Lilley said the amount of money to be made by developing land in a burgeoning Asian economy "There is not as much money generated in leaving a conservation area as it is...as there is in building big roads, airstrips, big hotels, swimming pools...," he said.

Motivation

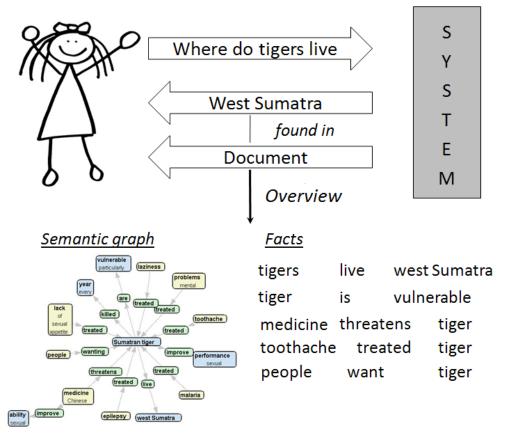
FEATUR

Two of Indonesia's According to tradit "Poaching is chan Lilley said tigers in "This is...perhaps According to South That's a lot of tiger Speaking this wee "Our estimate is th Extinction of the tic Despite concerns "Tiger populations The WWF estimat Subijanto, a spoke The Kerinci conse Whilse the Sumat

"There is not as m



System Overview

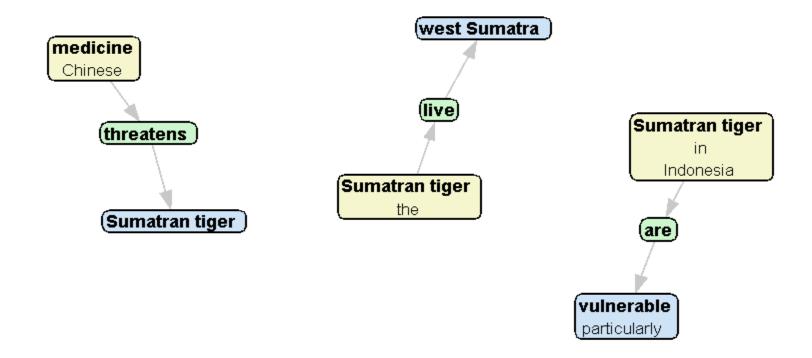


Summary

There are people wanting tiger products who didn't want them before, "Ron Lilley, coordinator for species conservation at the WWF in Jakarta, told Reuters. Subijanto, a spokesman for the Forestry Ministry, said Indonesia was committed to protecting the tigers, which live within Sumatra's four designated conservation areas.

Triplets

- Facts stated in the text
- The core of the sentence (subject, verb, object)

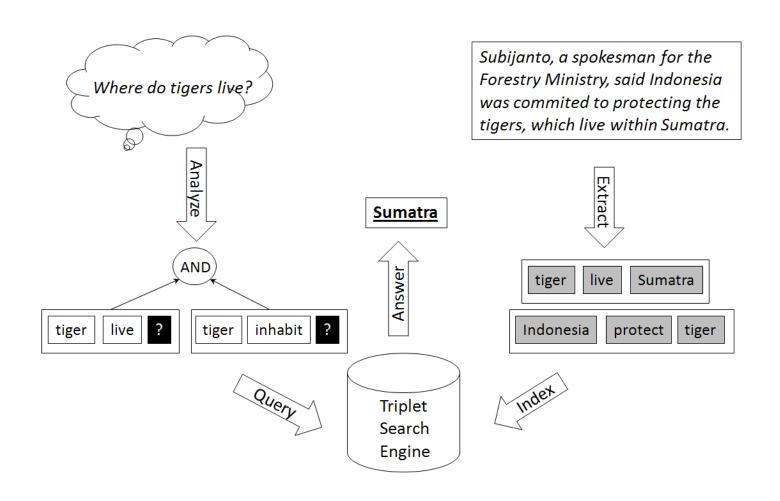


Question Answering

Offline:

- Extract facts (triplets) from text
- Index triplets to enable structured search on them Online:
- Analyze questions to obtain the queries for the triplet search
- Retrieve the answer and the document containing it
- Browse the document overview

Question Answering



Question Types

Type

- Yes/No questions
- list questions
- quantity questions
- location questions
- time questions
- reason questions

Example

- Do animals eat fruit?
- What do animals eat?
- How much fruit do animals eat?
- Where do animals eat?
- When do animals eat?
- Why do animals eat fruit?

Parsing → treebank → query expansion (WordNet)



where do tigers live

Ask

We found that

live	the following
surviving	world
live	Sumatra
	surviving



world CHINA: FEATURE - Tigers must earn their meat in China. With only about 300 Siberian tigers surviving in the world, and only 20 in the wild in China, that help must come soon, said Liu.

Sumatra INDONESIA: FEATURE -

Chinese medicine threatens Sumatran tiger. Subijanto, a spokesman for the Forestry Ministry, said Indonesia was committed to protecting the tigers, which live within Sumatra's four designated conservation areas.



Enter a question or topic

Ask



INDONESIA: FEATURE - Chinese medicine threatens Sumatran tiger.

Show Document Overview

FEATURE - Chinese medicine threatens Sumatran tiger.

Two of Indonesia's three species of tigers have become extinct and the third is severely threated because of the belief that tiger parts can improve sexual performance, experts say.

According to traditional Chinese medical belief, mental problems, laziness, malaria, epilepsy, toothache and lack of sexual appetite can be treated with tiger parts, leading to rampant poaching of the animal in Asia, the World Wide Fund for Nature (WWF) said.

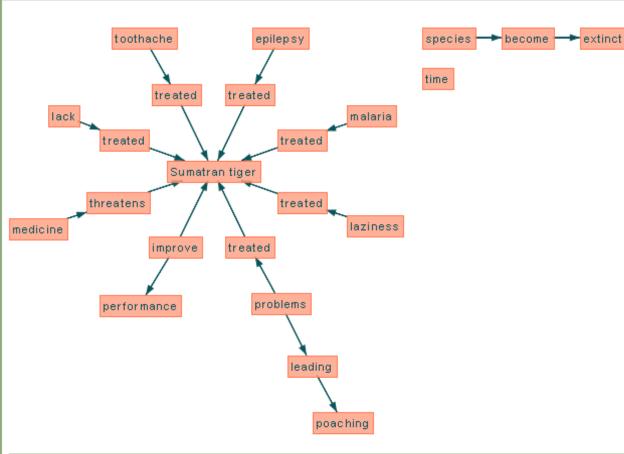
"Poaching is changing all the time...this trend for tiger products is new. There are people wanting tiger products who didn't want them before," Ron Lilley, coordinator for species conservation at the VWVF in Jakarta, told Reuters.

Lilley said tigers in Indonesia, now reduced to about 500 animals on the island of Sumatra, are particularly vulnerable. Poaching is a severe threat, with huge markets for tiger bones in China, South Korea and Taiwan.

"This is...perhaps a billion-dollar industry, mostly illegal and run by people not interested in sustainability," he said.

According to South Korean immigration statistics, the country imported 3,994 kilograms (8787 pounds) of tiger bones from Indonesia between 1970 and 1993, making Indonesia their largest supplier.

That's a lot of tigers, Lilley said, since the bones of one tiger weigh less than 10 kilograms (22 pounds).



SACTS medicine threatens Sumatran tiger species become extinct tiger parts improve performance problems treated tiger parts problems leading poaching laziness treated tiger parts malaria treated tiger parts epilepsy treated tiger parts toothache treated tiger parts lack treated tiger parts Poaching changing time



Two of Indonesia's three species of tigers have become extinct and the third is severely threated because of the belief that tiger parts can improve sexual performance, experts say. According to traditional Chinese medical belief, mental problems, laziness, malaria, epilepsy, toothache and lack of sexual appetite can be treated with tiger parts, leading to rampant poaching of the animal in Asia, the World Wide Fund for Nature (WWF) said. "Poaching is changing all the time... this trend for tiger products is new.

Document Overview

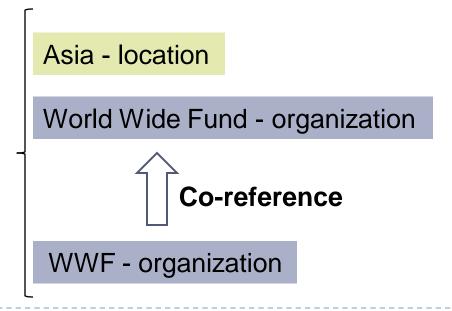
Analyze the document containing the answer:

- Highlight facts described by subject verb object triplets (identified in the Penn Treebank parse tree)
- Obtain the document semantic graph
- View the automatic document summary

FEATURE - Chinese medicine threatens Sumatran tiger. The displacement of the chinese of the chi



Named entity extraction Co-reference resolution According to traditional Chinese medical belief, mental problems, laziness, malaria, epilepsy, toothache and lack of sexual appetite can be treated with tiger parts, leading to rampant poaching of the animal in **Asia**, the **World Wide Fund** (**WWF**) said.



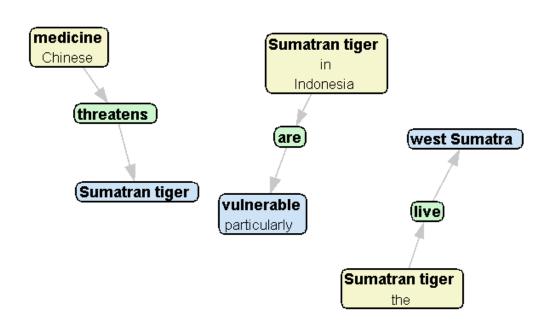




Named entity extraction Co-reference resolution



S - V - O triplet extraction





Named entity extraction Co-reference resolution

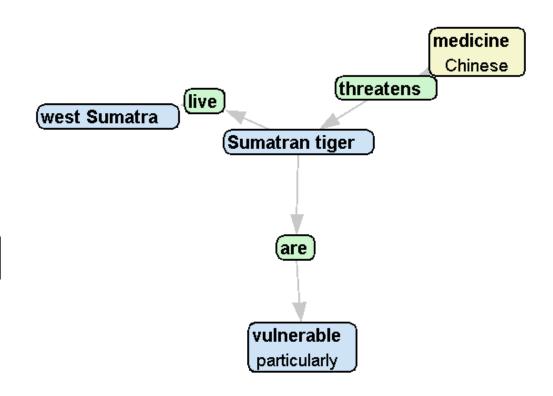


S - V - O triplet extraction

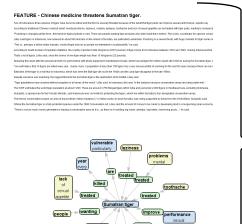


Triplet enhancement





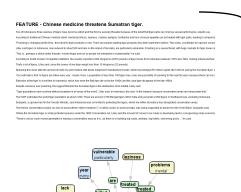
FEATURE - Chinese modicine threatens Sumatran tiger. Institution of the control of the control





Feature Extractor







Feature Extractor



Linear SVM

Features:

- linguistic
- document
- graph

FEATURE - Chinese medicine threatens Sumatran tiger. The distance has seen of grantes access and a feature than seen had been been seen and a feature than seen and a feature





Feature Extractor



Linear SVM

Linear Model



Ranking

The Kerinci conservation project, an area of around three million hectares (7. 4 million acres) in west Sumatra, was being supported by funds from the World Bank, Subijanto said. [10.0912] Subijanto, a spokesman for the Forestry Ministry, said Indonesia was committed to protecting the tigers, which live within Sumatra's four designated conservation areas. [9.4155]

FEATURE - Chinese medicine threatens Sumatran tiger.

Two of Indonesia's three species of tigers have become extinct and the third is severely threated because of the belief that tiger parts can improve sexual performance, experts say.

According to traditional Chinese medical belief, mental problems, laziness, malaria, epilepsy, toothache and lack of sexual appetite can be treated with tiger parts, leading to rampant problems, laziness, malaria, epilepsy, toothache and lack of sexual appetite can be treated with tiger parts, leading to rampant proposition in the control of tiger products in the part of tiger parts, leading to rampant proposition in the control of tiger products in the part of tiger parts, leading to rampant proposition in the control of tiger parts, leading to rampant proposition in the part of tiger parts can improve sexual performance, experts say.

According to traditional Chinese medical belief, mental problems, laziness, malaria, epilepsy, toothache and lack of sexual appetite can be treated with tiger parts, leading to rampant proposition in the control of the parts of tiger parts and lack of sexual appetite can be treated with tiger parts, leading to rampant proposition in the control of the parts of the parts of the parts of tiger parts of the part

According to South Korean immigration statistics, the country imported 3,994 kilograms (8787 pounds) of tiger bones from Indonesia between 1970 and 1993, making Indonesia their That's a lot of tigers, Lilley said, since the bones of one tiger weigh less than 10 kilograms (22 pounds).

Speaking this week after the announcement of a joint venture with photo equipment manufacturer Kodak, which has pledged 50 million rupiah (\$21,000) to saving the Sumatran tiger, L "Our estimate is that 14 tigers are killed every year...maybe more. A population of less than 100 tigers has a very low possibility of surviving for the next 50 years because there can be n Extinction of the tiger is a real fear in Indonesia, which has seen the Bali tiger die out In the 1940s and the Java tiger disappear in the late 1980s.

Despite concerns over poaching, the biggest threat to the Sumatran tiger is the destruction of its habitat, Lilley said.

"Tiger populations have crashed without exception in all areas of the world," Lilley said. In Indonesia, this was "in the balance because conservation areas are being eaten into."

The WWF estimates the world tiger population at about 7,200. There are around 3,750 Bengal tigers left in India and just under 2,000 tigers in Southeast Asia, excluding Indonesia.

Subijanto, a spokesman for the Forestry Ministry, said Indonesia was committed to protecting the tigers, which live within Sumatra's four designated conservation areas.

The Kerinci conservation project, an area of around three million hectares (7.4 million acres) in west Sumatra, was being supported by funds from the World Bank, Subijanto said.

Whilse the Sumatran tiger is a fully protected species under the 1990 Conservation Act, Lilley said the amount of money to be made by developing land in a burgeoning Asian economy

"There is not as much money generated in leaving a conservation area as it is...as there is in building big roads, airstrips, big hotels, swimming pools...," he said.

FEATURE - Chinese medicine threatens Sumatran tiger.

Two of Indonesia's three species of tigers have become extinct and the third is severely threated because of the belief that tiger parts can improve sexual performance, experts say.

There are people wanting tiger products who didn't want them cies consei Lilley s er bones in before, "Ron Lilley, coordinator for species conservation at the "This WWF in Jakarta, told Reuters. nesia their Subijanto, a spokesman for the Forestry Ministry, said Speak tran tiger, L Indonesia was committed to protecting the tigers, which live re can be n within Sumatra's four designated conservation areas. Extincti Despi The Kerinci conservation project, an area of around three "Tiger million hectares (7. 4 million acres) in west Sumatra, was onesia being supported by funds from the World Bank, Subijanto said.

Whilse the Sumatran tiger is a fully protected species under the 1990 Conservation Act, Lilley said the amount of money to be made by developing land in a burgeoning Asian economy

[&]quot;There is not as much money generated in leaving a conservation area as it is...as there is in building big roads, airstrips, big hotels, swimming pools...," he said.

Conclusions

- Enhanced question answering system
- Question answering, where the answer is supported by documents
- Document browsing
 - Facts
 - Document semantic graph
 - Automatic document summary

References

Wikipedia:

http://en.wikipedia.org/wiki/Question_answering

Question answering links:

http://alcor.concordia.ca/~smw/home/qa.html

- Recognizing Textual Entailment using Logical Inference \ Yaroslav Riabinin
- The Survey of Knowledge based Question Answering Systems
 Lee Geun-hae, Kristofer Lofgren
- TREC 2007 QA Data:

http://trec.nist.gov/data/qa/t2007_qadata.html

Question Answering based on semantic graphs:

http://km.aifb.uni-karlsruhe.de/ws/semsearch09/semse2009_22.pdf

The AnswerArt system homepage:

http://answerart.net/

Thanks for listening!

Questions?

Don't ask me, ask Jeeves...