



Probabilistic Question Answering on the Web

**Dragomir Radev, Weiguo Fan, Hong Qi, Harris Wu,
Amardeep Grewal**

**The University of Michigan
Ann Arbor, MI, 48109-1092**

{radev, wfan, hqi, harriswu, agrewal}@umich.edu

<http://perun.si.umich.edu/clair>

People ask questions

- Excite corpus of 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 set up TCP/IP) or other
- In other words, 100 K questions per day

People ask questions

In what year did baseball become an official sport?

Who is the largest man in the world?

Where can I get information on Raphael?

where can I find information on puritan religion?

Where can I find how much my house is worth?

how do I get out of debt?

Where can I find 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 fountain pen?

how do I set up TCP/IP?

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?

What is the largest city in Northern Afghanistan?

Welcome!

Crisis Update:



Editions: Europe | Asia

1000 Hours FREE for 45 days

- TRY AOL FREE!
- TRY AOL FREE!
- TRY AOL FREE!
- TRY AOL FREE!

Updated: 08:07 a.m. EST (1307 GMT) -- 10 November 2001

SEARCH GO

MAIN PAGE

- WORLD
- U.S.
- WEATHER
- BUSINESS
- SPORTS
- POLITICS
- LAW
- SCI-TECH
- SPACE
- HEALTH
- ENTERTAINMENT
- TRAVEL
- EDUCATION
- CAREER
- IN-DEPTH

- QUICK NEWS
- LOCAL
- COMMUNITY
- MULTIMEDIA
- E-MAIL SERVICES
- CNNtoGO
- ABOUT US

- CNN TV
- what's on
- show transcripts
- CNN Headline News
- CNN International
- askCNN

WAR AGAINST TERROR



(AP PHOTO)

A Northern Alliance sniper in Northern Afghanistan

Taliban: Key city has fallen

- The Taliban confirm Mazar-e Sharif has been taken by the Northern Alliance.
- Pakistan newspaper: Al Qaeda has nuclear and chemical weapons
- President Bush to address U.N. today
- French authorities detain five people suspected of planning terrorist attacks

[FULL STORY >>](#)

Why is Mazar-e Sharif significant?

TIME.com: [Rebels: Mazar-e Sharif is ours](#)

Military Desk Briefing | War Room

Afghanistan: [Info](#) | [Strikes map](#)

Voices: [Message board and chats](#)

ON THE SCENE

[more>>](#)

Kamal Hyder: ['Very rough night'](#)

Kelly Wallace: [Anti-terrorism strategy](#)

FRONT LINES

[more>>](#)

- [Bin Laden makes nuclear threat](#)
- Fact Sheet: [The battle for Mazar-e Sharif](#)

AMERICA AT HOME

[more>>](#)

- [Bush to seek coalition support at U.N.](#)
- [N.Y. magazine office set for anthrax tests](#)
- [FBI profiles anthrax letter writer](#)
- [N.J. postal sites positive for anthrax](#)
- [Airport security company undergoes shakeup](#)
- [Christmas tree up at Rockefeller Center](#)
- Fact Sheet: [Phony anthrax threats](#)

OTHER TOP NEWS

- [China set to join WTO](#)
- [Russian market blast kills four](#)
- [Howard wins third term after tight race](#)
- [Climate treaty set to be ratified](#)
- [Rome rain dampens war tensions](#)

BUSINESS at [CNNmoney](#) [more >](#)

sponsored by:

- [Big IPO week on tap](#)
- [Dynergy to buy Enron for \\$9.5B](#)
- [GM recalls 86,000 SUVs](#)
- [Personal Finance](#)

Markets: 4:30pm ET, 11/9

stock quotes:

DJIA \uparrow 20.50

NAS \uparrow 0.71

S&P \uparrow 1.77

[Get Quote](#)

FOLLOW THE NEWS

[CNN to go](#) [WIDEWORLD NEWS DELIVERED](#) [Click here](#)

Question answering

What is the largest city in Northern Afghanistan?



Possible approaches

- Map?

- Knowledge base

Find x : $\text{city}(x) \wedge \text{located}(x, \text{"Northern Afghanistan"}) \wedge$
 $\wedge \neg \text{exists}(y): \text{city}(y) \wedge \text{located}(y, \text{"Northern Afghanistan"}) \wedge$
 $\wedge \text{greaterthan}(\text{population}(y), \text{population}(x))$

- Database?
- World factbook?
- Search engine?



User interfaces to the Web

- Command-line search interfaces
- speech/natural language
- Procedural vs. exact answers
- Ask Jeeves?



... **Afghanistan**, **Kabul**, 2,450 ... Administrative capital and **largest city** (1997 est ... Undetermined.
Panama, Panama **City**, 450,668. ... of the Gauteng, **Northern** Province, Mpumalanga ...
www.infoplease.com/cgi-bin/id/A0855603

... died in **Kano**, **northern** Nigeria's **largest city**, during two days of anti-American riots
led by Muslims protesting the US-led bombing of **Afghanistan**, according to ...
www.washingtonpost.com/wp-dyn/print/world/

... air strikes on the **city**. ... the Taliban militia in **northern Afghanistan** in a significant
blow ... defection would be the **largest** since the United States ...
www.afgha.com/index.php - 60k

... **Kabul** is the capital and **largest city** of **Afghanistan**. met. area pop. 2,029,889),
is the **largest city** in Uttar Pradesh, a state in **northern** India.
school.discovery.com/homeworkhelp/worldbook/atozgeography/k/k1menu.html

... **Gudermes**, Chechnya's second **largest** town. The attack ... location in **Afghanistan's** outlying
regions ... in the **city** of **Mazar-i-Sharif**, a **Northern** Alliance-affiliated ...
english.pravda.ru/hotspots/2001/09/17/

... Get Worse By RICK BRAGG Pakistan's **largest city** is getting a jump on the ... Region: Education
Offers Women in **Northern Afghanistan** a Ray of Hope. ...
www.nytimes.com/pages/world/asia/

... within three miles of the airport at **Mazar-e-Sharif**, the **largest city** in **northern**
Afghanistan, held since 1998 by the Taliban. There was no immediate comment ...
uk.fc.yahoo.com/photos/a/afghanistan.html



What is the largest city in Northern Afghanistan?

Query modulation

(largest OR biggest) city "Northern Afghanistan"

Document retrieval

www.infoplease.com/cgi-bin/id/A0855603
www.washingtonpost.com/wp-dyn/print/world/

Sentence retrieval

Gudermes, Chechnya's second **largest** town ... location in **Afghanistan's** outlying regions within three miles of the airport at Mazar-e-Sharif, the **largest city** in **northern Afghanistan**

Answer extraction

Gudermes
Mazer-e-Sharif

Answer ranking

Mazer-e-Sharif
Gudermes

Excite Search Results for "when did the Neanderthal man live?" - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites History Print Copy Paste Links >>

Address http://search.excite.com/search.gw?c=web&search=when+did+the+Neanderthal+man+live%3F Go

SIGN IN/SIGN UP FREE EMAIL CHAT NOW! PHOTO ALBUMS SEND E-CARDS Hot Screen Savers Browser Skins

ONLINE GAMES 

[Click Here!](#)
[Click Here!](#)

 **precision search** 

Search: Web Sites News Products Photos MP3/Audio Video

Quick Find [Excite](#) | [Help](#)


[Search Home](#) | [Advanced Web Search](#) | [Voyeur!](#) | [Submit A Site](#)

About 5,037,040 found in **Web Sites** for: **when did the Neanderthal man live?**

Show Results: The Web | [Web Directory](#)

Top 10 Displayed: [Full Description](#) | [Show Titles Only](#) | [View by URL](#)

- [Willkommen bei neanderthal.de](#)
URL: <http://www.neanderthal.de/>
Excite Directory Match: [Dusseldorf, Germany - Services and Activities](#)
Das neue Neanderthal Museum - ein interaktives Erlebnismuseum zur Entwicklungsgeschichte der Menschheit.
- [Neanderthal Flute](#)
URL: <http://www.webster.sk.ca/greenwich/fl-compl.htm>
Excite Directory Match: [Neanderthals](#)
Musicology analysis of oldest musical instrument showing it played notes of do, re, mi, scale. Significant to concepts of the origin of music.
- [Neandertals: A Cyber Perspective](#)
URL: <http://thunder.indstate.edu/~ramanank/>
Neandertals: A Cyber Perspective by Kharlena Maria Ramanan In 1856 at the Feldhofer Cave, near Dusseldorf Germany, Neanderthal Man informally introduced himself to the world.
- [Neanderthal: Neanderthal Man And The Story Of Human Origins: Jordan, P...](#)
URL: <http://www.netstoreusa.com/stbooks/075/0750919345.shtml>
Neanderthal: Neanderthal Man And The Story Of Human Origins by Jordan, Paul; English; Hardcover
- [R/P 38 Neanderthal = Nephilim? Carbon 14 dating?](#)
URL: <http://www.fttech.net/~bric/rp.no38.html>
Research Paper 38 Size 16 A4 pages. 106K Bytes. Written by John Denton 28th Dec 1996 A speculative research paper examining current evidence available on Neanderthal man with comparison to refer...
- [The Neanderthal Man's Cave](#)
URL: <http://www.rossiresort.com/guests/grandpa/>
I am Out Cold!! Send comments to curator@rossiresort.com Copyright © 1999 by The Rossi Resort. This page was last updated on 24 June 1999 at 09:28.
- [Early Man: Neanderthal Man](#)
URL: <http://mall.turnpike.net/C/cs/emnh.htm>
In 1856 workers quarrying for limestone in the Neander Valley near Dusseldorf, Germany came across a skull and bones. In the succeeding years many other specimens were found, not only in the Neander...
- [Neanderthal DNA Soup](#)
URL: http://www.ridgenet.net/~do_while/sage/v1i12n.htm
Evolution in the News - September 1997 The recent news reports about the relationship between Neanderthal man and modern man were based on a study by Matthias Kings, Anne Stone, Ralf W.


[WHEN DID T...](#)
[Buy Books Here](#)
[Search Amazon](#)

Submit Your Site!
 • Reach 83% of web users
 • Reviewed within 48 hours

[Click here](#)
[Click Here!](#)

More Excite
[Sexual Health](#)
[Photo Personals](#)
[Free Wallpaper](#)
[Music Radio](#)
[Free Screensaver](#)
[Send a Kiss](#)
[Business Tools](#)

[Click Here!](#) 

Excite Search Results for "live OR lived AND "Neanderthal man"" - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites History Print Copy Paste Links >>

Address http://search.excite.com/search.gw?c=web&showSummary=true&lang=en&sl=live+OR+lived+AND+%22Neanderthal+man%22 Go

SIGN IN/SIGN UP FREE EMAIL CHAT NOW! PHOTO ALBUMS SEND E-CARDS Hot Screen Savers Browser Skins

SDKIPFSODVVNCVNSGGCINFCNVSVJ
 VONLNFETBKMAKEAPLWQMVICKYE
 HTJABIYKFORLCWPFRIENDSYOERU
 SLOWNCAMESTUNEMQWCXPZALPDKG
 KWIORNMOOKLGPDLISYKHNOWQOMVN

excite

Click Here!
 Click Here!

excite precision search

Search: Web Sites News Products Photos MP3/Audio Video

Quick Find

live OR lived AND "Neanderthal man" Search

Excite | Help

Search Home | Advanced Web Search | Voyeur! | Submit A Site

About 7,540 found in **Web Sites** for: **live OR lived AND "Neanderthal man"**

Show Results: The Web | [Web Directory](#)

amazon.com

- [LIVE OR LI...](#)
- [Buy Books Here](#)
- [Search Amazon](#)

More Excite

- [Sexual Health](#)
- [Photo Personals](#)
- [Free Wallpaper](#)
- [Music Radio](#)
- [Free Screensaver](#)
- [Send a Kiss](#)
- [Business Tools](#)

Top 10 Displayed: Full Description | [Show Titles Only](#) | [View by URL](#)

- [THE EXTINCTION OF NEANDERTHAL MAN 10-3-95](#)
 URL: <http://sciboard.louisville.edu/news/scinews/bioeva/neander.htm>
 About 40,000 years ago, Cro-Magnon man ("modern" man) lived in Africa and the Middle East, while Europe was the home of Neanderthals. Both were Homo sapiens (meaning they could probably ...
- [CNN - DNA tests show humans not Neanderthals' descendants - December 1...](#)
 URL: <http://www.cnn.com/TECH/9712/17/britain.neanderthals/index.html>
 Neanderthal man From Correspondent Siobhan Darrow LONDON (CNN) -- For more than a century, scientists believed that Neanderthal man, a hulking ape-like creature who lived in Europe and the Middle...
- [Eyring-L: Humans not Neanderthals' descendants](#)
 URL: <http://cyberbuzz.gatech.edu/kaboom/mail-archives/Eyring-L/97...>
 DNA tests show humans not Neanderthals' descendants Neanderthal man December 17, 1997 Web posted at: 11:11 p.m. EST (0411 GMT) >From Correspondent Siobhan Darrow LONDON (CNN) -- For more than a...
- [sci.anthropology.paleo: Sep-95: DISCOVER/Neanderthal/Homo Sap.](#)
 URL: <http://www.anatomy.usyd.edu.au/danny/anthropology/sci.anthro...>
 Peggy Hall (phall@primenet.com) 1 Sep 1995 05:11:24 GMT I just picked up the new Discover Magazine. There is a question posed on the cover regarding Neanderthals and Homo Sapiens living side by s...
- [sci.anthropology.paleo: Sep-95: Re: DISCOVER/Neanderthal/Homo Sap.](#)
 URL: <http://www.anatomy.usyd.edu.au/danny/anthropology/sci.anthro...>
 In article <427hb\$len@huitzilo.tezcat.com>, ermiller@tezcat.com (Erin Miller) writes: >> >Funny, I live side by side with *lots* of species I don't have sex with. > >-erin
- [sci.anthropology.paleo: Sep-95: Re: DISCOVER/Neanderthal/Homo Sap.](#)
 URL: <http://www.anatomy.usyd.edu.au/danny/anthropology/sci.anthro...>
 Peggy Hall wrote: I just picked up the new Discover Magazine. There is a question posed on the cover regarding Neanderthals and Homo Sapiens living side by side for so long: How could we live sid...
- [Re: Hubbys/Bears and No Sleep](#)
 URL: <http://firlist.sparklist.com/9810/msg07560.html>
 Subject: Re: Hubbys/Bears and No Sleep From: Linda Coco <lccoco@cyberport.net> I live in Bedrock with Neanderthal man. We're a modern stone age family.
- [Re: Hubbys/Bears and No Sleep](#)
 URL: <http://firlist.sparklist.com/9810/msg07734.html>
 Subject: Re: Hubbys/Bears and No Sleep From: Melissa O <mellie_27@yahoo.com> Lind-lu..I'll bet it is beautiful there=) Do you like it? Well, stay safe=) melissa=) ---Linda Coco <lccoco@cyb...

Click Here! Internet

Research problems

- Source identification:
 - semi-structured vs. text sources
- Query modulation:
 - best paraphrase of a NL question given the syntax of a search engine?
 - Compare two approaches: noisy channel model and rule-based
- Sentence ranking
 - n-gram matching, Okapi, co-reference?
- Answer extraction
 - question type identification
 - phrase chunking
 - no general-purpose named entity tagger available
- Answer ranking
 - what are the best predictors of a phrase being the answer to a given question: question type, proximity to query words, frequency
- Evaluation (MRDR)
 - accuracy, reliability, timeliness



Document retrieval

- Use existing search engines: Google, AlltheWeb, NorthernLight
- No modifications to question
- CF: work on QASM (ACM CIKM 2001)



Sentence ranking

- Weighted N-gram matching:

$$S_i = \frac{w_1 * \sum_{i=1}^{N_1} tf_i * idf_i + w_2 * \sum_{j=1}^{N_2} tf_j + w_3 * \sum_{k=1}^{N_3} tf_k}{F}$$

- Weights are determined empirically, e.g., 0.6, 0.3, and 0.1

Probabilistic phrase reranking

- Answer extraction: probabilistic phrase reranking. What is:

$$p(ph \text{ is answer to } q \mid q, ph)$$

- Evaluation: TRDR

- Example: (2,8,10) gives $.725 \frac{1}{n} \left(\sum_{i=1}^n \frac{1}{r_i} \right)$
- Document, sentence, or phrase level

- Criterion: presence of answer(s)
- High correlation with manual assessment

Phrase types

PERSON PLACE DATE NUMBER DEFINITION
ORGANIZATION DESCRIPTION ABBREVIATION
KNOWNFOR RATE LENGTH MONEY REASON
DURATION PURPOSE NOMINAL OTHER

Question Type Identification

- Wh-type not sufficient:
 - Who: PERSON 77, DESCRIPTION 19, ORG 6
 - What: NOMINAL 78, PLACE 27, DEF26, PERSON 18, ORG 16, NUMBER 14, etc.
 - How: NUMBER 33, LENGTH 6, RATE 2, etc.
- Ripper:
 - 13 features: Question-Words, Wh-Word, Word-Beside-Wh-Word, Is-Noun-Length, Is-Noun-Person, etc.
 - Top 2 question types
- Heuristic algorithm:
 - About 100 regular expressions based on words and parts of speech

Ripper performance

Training	Test	Train Error Rate	Test Error Rate
TREC9	TREC8	22.4%	24%
TREC8,9	TREC10	17.03%	30%
TREC8,9,10	-	20.69%	-

Regex performance

Training	Test on TREC9	Test on TREC8	Test on TREC10
TREC9	7.8%	15%	18%
TREC8,9	7.4%	6%	18.2%
TREC8,9,10	4.6%	5.5%	7.6%



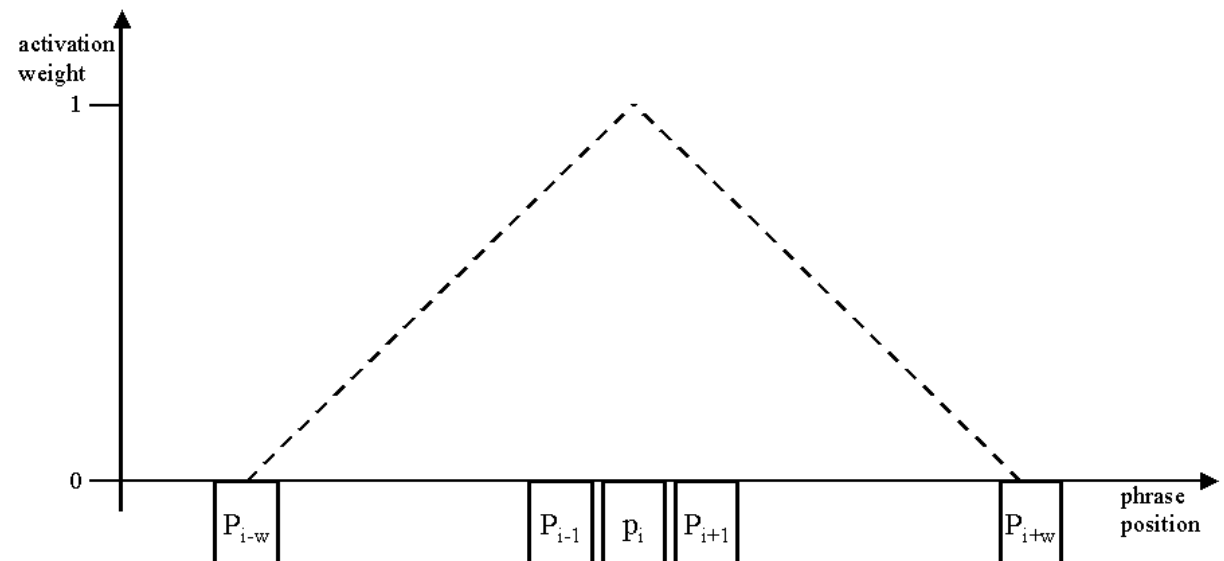
Phrase ranking

- Phrases are identified by a shallow parser (Itchunk from Edinburgh)
- Four features:
 - Proximity
 - POS (part-of-speech) signature (qtype)
 - Query overlap
 - Frequency



Proximity

- Phrasal answers tend to appear near words from the query
- Average distance = 7 words, range = 1 to 50 words
- Use linear rescaling of scores



Part of speech signature

Penn Treebank tagset (DT = determiner, JJ = adjective)

Signature	Phrase Types
VBD	NO (100%)
DT NN	NO (86.7%) PERSON (3.8%) NUMBER (3.8%) ORG (2.5%)
NNP	PERSON (37.4%) PLACE (29.6%) DATE (21.7%) NO (7.6%)
DT JJ NNP	NO (75.6%) NUMBER (11.1%) PLACE (4.4%) ORG (4.4%)
NNP NNP	PLACE (37.3%) PERSON (35.6%) NO (16.9%) ORG (10.2%)
DT NNP	ORG (55.6%) NO (33.3%) PLACE (5.6%) DATE (5.6%)

Example: “Hugo/NNP Young/NNP”

$$P(\text{PERSON} \mid \text{“NNP NNP”}) = .458$$

Example: “the/DT Space/NNP Flight/NNP Operations/NNP contractor/NN”

$$P(\text{PERSON} \mid \text{“DT NNP NNP NNP NN”}) = 0$$

Query overlap and frequency

- Query overlap:
 - What is the capital of Zimbabwe?
 - Possible choices:
Mugabe, Zimbabwe, Luanda, Harare
- Frequency:
 - Not necessarily accurate but rather useful



Reranking

Rank	Probability and phrase
1	0.599862 the_DT Space_NNP Flight_NNP Operations_NNP contractor_NN ._.
2	0.598564 International_NNP Space_NNP Station_NNP Alpha_NNP
3	0.598398 International_NNP Space_NNP Station_NNP
4	0.598125 to_TO become_VB
5	0.594763 a_DT joint_JJ venture_NN United_NNP Space_NNP Alliance_NNP
6	0.593933 NASA_NNP Johnson_NNP Space_NNP Center_NNP
7	0.587140 will_MD form_VB
8	0.585410 The_DT purpose_NN
9	0.576797 prime_JJ contracts_NNS
10	0.568013 First_NNP American_NNP
11	0.567361 this_DT bulletin_NN board_NN
12	0.565757 Space_NNP :_:
13	0.562627 'Spirit_NN ' _" of_IN
14	0.561702 space_NN
15	0.561064 February_NPN
...	
41	0.516368 Alan_NNP Shepard_NNP

Proximity = .5164



Reranking

Rank	Probability and phrase
1	0.465012 Space_NNP Administration_NNP ._.
2	0.446466 SPACE_NNP CALENDAR_NNP _.
3	0.413976 First_NNP American_NNP
4	0.399043 International_NNP Space_NNP Station_NNP Alpha_NNP
5	0.396250 her_PRP\$ third_JJ space_NN mission_NN
6	0.395956 NASA_NNP Johnson_NNP Space_NNP Center_NNP
7	0.394122 the_DT American_NNP Commercial_NNP Launch_NNP Industry_NNP
8	0.390163 the_DT Red_NNP Planet_NNP ._.
9	0.379797 First_NNP American_NNP
10	0.376336 Alan_NNP Shepard_NNP
11	0.375669 February_NNP
12	0.374813 Space_NNP
13	0.373999 International_NNP Space_NNP Station_NNP
14	0.372289 Als_NNPS
15	0.371194 The_NNP Spirit_NNP

Qtype = .7288

Proximity * qtype = .3763



Reranking

Rank	Probability and phrase
1	0.478857 Neptune_NNP Beach_NNP ._.
2	0.449232 February_NNP
3	0.447075 Go_NNP
4	0.437895 Space_NNP
5	0.431835 Go_NNP
6	0.424678 Alan_NNP Shepard_NNP
7	0.423855 First_NNP American_NNP
8	0.421133 Space_NNP May_NNP
9	0.411065 First_NNP American_NNP woman_NN
10	0.401994 Life_NNP Sciences_NNP
11	0.385763 Space_NNP Shuttle_NNP Discovery_NNP STS-60_NN
12	0.381865 the_DT Moon_NNP International_NNP Space_NNP Station_NNP
13	0.370030 Space_NNP Research_NNP A_NNP Session_NNP
14	0.366714 First_NNP American_NNP
15	0.359058 Sally_NNP Ride_NNP Sally_NNP Ride_NNP

All four features

NSIR: A Web-Based Question Answering System

Your question: (TREC-8 TREC-9)

Answer: (for evaluation) [Tip](#)

Engine: Return: No. of documents: No. of answers: Debug

[Help](#)

Running 'Who was the first American in space?' on Yahoo.

DOCUMENTS

- [1] <http://www2.lucidcafe.com/lucidcafe/library/96may/ride.html>: Sally Kristen Ride | First American Woman in Space
Sally Kristen Ride First American Woman in Space Resources Menu Categorical Index Caf Library Gall
- [2] <http://www.hq.nasa.gov/office/pao/History/Timeline/100ft.html>: Chronology of Selected Highlights in the First 100 American ...
Idoctype html public -//IETF//DTD HTML//EN National Aeronautics and Space Administration Chronolo
- [3] http://www.space.com/news/spacehistory/shepard_anniversary_010504-1.html: Alan Shepard: The First American Astronaut
Alan Shepard: The First American Astronaut Thurs. May 9 2002 News Missions/Launches Science/Astronom
- [4] http://neurolab.jsc.nasa.gov/trivia_pers.htm: NASA Neurolab Web: Space Trivia Page
NASA Neurolab Web: Space Trivia Page People Firsts Women in Space Age Other Firsts Who was the fi
- [5] <http://www.enc.org/thisweek/calendar/unit0.1819.51.00.shtml>: ENC : This Week : Classroom Calendar : First American in Space
ENC : This Week : Classroom Calendar : First American in Space YOU ARE HERE ENC Home This Week Cl
- [6] http://news.bbc.co.uk/1/hi/english/sci/tech/newsid_137000/137462.stm: BBC News | Sci/Tech | First American in space dies

Question type guesses: PERSON DESCRIPTION POSSIBLE ANSWERS

order	phrase(score)	proxi	sig_prob	q_olap	freq	pos
1	the first american (55.689)	6.3032	0.500000	1.00	16	DT JJ NNP
2	science (49.367)	7.9563	0.364407	1.00	8	NNP
3	the first american woman (47.415)	8.9810	0.250000	1.00	10	DT JJ JJ NN
4	may (43.641)	7.4807	0.364407	1.00	8	NNP
5	sally ride first american woman (33.780)	9.1275	1.000000	1.00	1	NNP NNP NNP NNP NN
6	the 20th century (33.391)	8.5372	0.500000	1.00	2	DT JJ NNP
7	december (30.081)	7.4083	0.364407	1.00	4	NNP
8	shepard (29.898)	6.3812	0.364407	1.00	7	NNP
9	july (28.758)	6.4932	0.364407	1.00	6	NNP

NSIR: A Web-Based Question Answering System

Your question: [TREC-8](#) [TREC-9](#) Answer: (for evaluation) [Tip](#)

When did French revolutionaries storm the Bastille? 1789(?;jul14)?July/s*14|

Engine: Return: No. of documents: No. of answers: Debug

[Help](#)

Running 'When did French revolutionaries storm the Bastille?' on Yahoo.

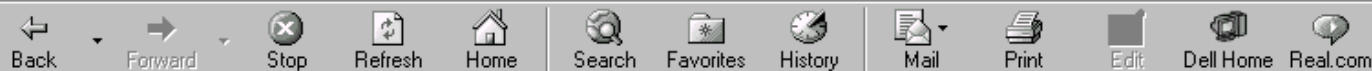
DOCUMENTS

- [1] <http://www.edunetconnect.com/cat/timemachine/250y.html>: [Baxter's EduNET - Time Machine](#)
Baxter's EduNET - Time Machine 250-300 years ago... The more things change the more they are the sam
- [2] <http://www.globalgourmet.com/destinations/france/bastille.html>: [France: Vive La Revolution](#)
- [3] <http://l2r.cs.uiuc.edu/~cogcomp/Data/QA/Trec8questions.txt>
<http://l2r.cs.uiuc.edu/~cogcomp/Data/QA/Trec8questions.txt>
A The Test Set of 200 Questions 1. Who is the author of the book ``The Iron Lady: A Biography of Ma
- [4] [http://teachers.lps.org/rraatz/stories/storyReader\\$34](http://teachers.lps.org/rraatz/stories/storyReader$34): [TeacherWeb | Russ Raatz: Chapter 22 - The French Revolution](#)
TeacherWeb Russ Raatz: Chapter 22 - The French Revolution Russ Raatz

Question type guesses: DATE

POSSIBLE ANSWERS

order	phrase (score)	proxi	sig_prob	q_olap	freq	pos
1	1789 while visiting (7.654)	4.3447	1.000000	1.00	1	CD IN VBG
2	22 (6.137)	3.3631	0.230769	1.00	10	CD
3	paris (5.729)	3.9826	0.211864	1.00	5	NNP
4	6961 (5.462)	7.6403	0.230769	1.00	1	CD
5	france (5.364)	3.5447	0.211864	1.00	7	NNP
6	161 (4.564)	6.9839	0.230769	1.00	1	CD
7	10 (4.551)	3.4013	0.230769	1.00	5	CD



[Evangelicalism in the Age of the ...](#)

'True' Christianity: English Evangelicalism in the Age of the French Revolution
'True' Christianity

[7] <http://www.schuykillriver.org/07-july/july-content2.htm>: [Untitled Document](#)

Untitled Document ongoing 9 Summer Family Concert. Lower Perkiomen Valley Park. 5-7 pm

12	california (3,344)	6.2393	0.211864	1.00	1	NNP
13	1 (2,772)	2.5333	0.230769	1.00	6	CD
14	this count clarkson (2,478)	3.4964	0.500000	1.00	1	DT NN NNP
15	162 (1,962)	4.5797	0.230769	1.00	1	CD
16	9 (1,801)	2.4957	0.230769	1.00	3	CD
17	europe (1,652)	2.2769	0.211864	1.00	4	NNP
18	america (1,622)	4.3447	0.211864	1.00	1	NNP
19	in 1789 (1,451)	4.0127	0.222222	1.00	1	IN CD
20	england (1,320)	1.9120	0.211864	1.00	5	NNP
<hr/>						
80	26 1789 (0,073)	0.6000	0.500000	1.00	1	CD CD
132	1789 (0,015)	0.4000	0.230769	1.00	1	CD
474	that in 1789 there (0,000)	1.1333	0.000000	1.00	1	IN IN CD EX

[FARR:0](#) [FHS:0](#) [NTRDR:1.053](#) [TRDR:1.075](#) [PREC:0.004](#) [MRR:0.000](#) [TRWR:1.060](#)
[NTRWR:1.043](#)

Document level performance

TREC 8 corpus (200 questions)

Engine	AlltheWeb	NLight	Google
Avg	0.8355	1.0495	1.3361
#>0	149	163	164

Sentence level performance

Engine	AW U	AW L	AW O	NL U	NL L	NL O	GO U	GO L	GO O
Avg	2.13	0.31	0.26	2.53	0.48	0.44	2.55	0.54	0.49
#>0	148	99	99	159	121	119	159	137	135

Phrase level performance

	AlltheWeb	NorthernLight	Google D+P	Google S+P
Upperbound	2.176	2.652	2.698	1.941
Appearance order	0.026	0.048	0.068	0.0646
Global proximity	0.038	0.054	0.058	0.0646
Combined	0.105	0.117	0.157	0.199

Experiments performed
 Oct-Nov. 2001

Discussion

- Questions+answers from competitors
- Google's limitations: number of words, API
- NorthernLight



Future work

- Further develop natural language interfaces to the Web
- Integration with our news summarization system NewsInEssence (<http://www.newsinessence.com>)
- Use question answering as a common front end for different applications (including access to semi-structured data)
- Semantic Web?

Conclusion

- Let the major search engines do what they are best at.
- Use Natural Language technology but ... to the extent feasible
- Deep parsing (e.g., Collins or Charniak parsers) is quite expensive [Kwok et al. 2001]
- Ignoring NLP is a bad idea



The CLAIR group

- The Computational Linguistics And Information Retrieval group at the University of Michigan
- <http://perun.si.umich.edu/clair>

The TREC Q&A evaluation

- Run by NIST [Voorhees and Tice 2000]
- 2GB of input
- 200 questions
- Essentially fact extraction
 - Who was Lincoln's secretary of state?
 - What does the Peugeot company manufacture?
- Questions are based on text
- Answers are assumed to be present
- No inference needed

nsir - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites History Print Copy Paste

Links Customize Links Free Hotmail Windows Liquid Music Network Welcome to Liquid Audio

Address n.si.umich.edu/~radev/nsir/nsir_test.cgi?question=Who+was+the+first+American+in+space%3F&engine=NorthernLight&return=answers&hits=10&Proceed=Proceed&sid=20011204143749 Go

Google Search Web Search Site PageRank Page Info Up Highlight

NSIR Demo

Question:

Engine: Return type: Number of hits:

Running 'Who was the first American in space?' on NorthernLight.

[1]
[PhysicsWeb - First American in space dies](#)
[PhysicsWeb - First American in space dies](#) (cached)
(<http://physicsweb.org/article/news/02/7/12>)
PhysicsWeb - First American in space dies Advanced site search news 2001 2000 1999 1998 1997 January February March April May June July August September October November December Show summaries quick search Search the news archive. Previous News for July 1998 Next First American in space dies 22 July 1998 Alan Shepard the first American to fly in space and one of only 12

[2]
[BBC News | Sci/Tech | First American in space dies](#)
[BBC News | Sci/Tech | First American in space dies](#) (cached)
(http://news.bbc.co.uk/1/hi/english/sci/tech/newsid_137000/137462.stm)
BBC News Sci/Tech First American in space dies Front Page World UK UK Politics Business Sci/Tech Health Education Sport Entertainment Talking Point On Air Feedback Low Graphics Help Wednesday July 22 1998 Published at 16:55 GMT 17:55 UK Sci/Tech First American in space dies Waiting for his first flight into space -1961 Alan Shepard the first American in space and the 5th man to walk upon the Mo

[3]
[Alan Shepard, First American in Space, Dies at 74](#)
[Alan Shepard, First American in Space, Dies at 74](#) (cached)
(<http://teachspace.org/news/latimes/shepard.html>)
Alan Shepard First American in Space Dies at 74 Back to Story This article appeared in the Los Angeles Times and is reprinted with permission Alan Shepard First American in Space Dies at 74 WASHINGTON -- Astronaut Alan Shepard the first American to fly in space and one of only 12 Americans to walk on the moon died late Tuesday at a hospital in California. He was 74. Alan Shepard

[4]
[NewsNet5 - First American In Space Dies](#)
[NewsNet5 - First American In Space Dies](#) (cached)
(<http://www.newsnet5.com/news/stories/news-980722-101353.html>)

Internet

nsir - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Forward Stop Home Search Favorites History Print Mail News

Links Customize Links Free Hotmail Windows Liquid Music Network Welcome to Liquid Audio

Address n.si.umich.edu/~radev/nsir/nsir_test.cgi?question=Who+was+the+first+American+in+space%3F&engine=NorthernLight&return=answers&hits=10&Proceed=Proceed&sid=20011204143749 Go

Google Search Web Search Site PageRank Page Info Up Highlight

[7]
[Alan Shepard, First American in Space, Dies at 74 -- 07/23/1998](#)
[Alan Shepard, First American in Space, Dies at 74 -- 07/23/1998](#) (cached)
(<http://www.cnsnews.org/indepth/archive/199807/IND19980723f.html>)
Alan Shepard First American in Space Dies at 74 -- 07/23/1998 Alan Shepard First American in Space Dies at 74 23 July 1998 By Lawrence Morahan CNS Staff Writer CNS Alan Shepard the first American in space and the fifth to walk on the moon died late Tuesday in a hospital in California. He was 74 and had suffered from leukemia. Those of us who are old enough to remember the early spaceflights wil

[9]
[Spirit' of First American In Space Remembered](#)
[Spirit' of First American In Space Remembered](#) (cached)
(http://ww.space.com/news/spacehistory/shepard_statue_000321.html)

[10]
[Spirit' of First American In Space Remembered](#)
[Spirit' of First American In Space Remembered](#) (cached)
(http://www.spacesight.com/news/spacehistory/shepard_statue_000321.html)

- 2.989800 Previous_JJ News_NNP
- 2.981519 Space_NNP Dies_NNP
- 2.953502 MoneyBackGuarantee_NNP Title_NN
- 2.571537 Space_NNP Dies_NNP
- 2.570339 KOIN_NNP
- 2.266675 July_NNP
- 2.248877 Waiting_NNP
- 2.202086 Space_NNP Dies_NNP
- 2.202086 MONTEREY_NNP Calif._NNP
- 2.202086 Alan_NNP Shepard_NNP
- 2.038934 KOIN_NNP
- 1.919728 July_NNP
- 1.888300 first_JJ American_NNP
- 1.822887 the_DT Space_NNP Shuttle_NNP
- 1.746002 Reuters_NNP
- 1.657909 the_DT Soviet_NNP Union_NNP
- 1.600629 Wednesday_NNP July_NNP
- 1.531870 Space_NNP Dies_NNP
- 1.504873 First_NNP American_NNP
- 1.497000 Cosmonaut_NNP Yuri_NNP Gagarin_NNP

Internet