

TEXAS
CONVERGENT

Concept Lecture: Natural Language Processing





What makes NLP so hard?

Ambiguity

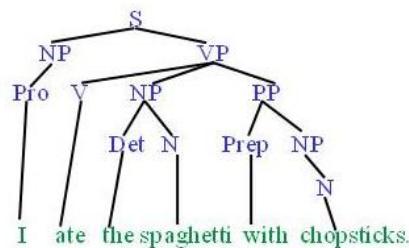
- *I saw the man on the hill with a telescope.*
- *I saw the Grand Canyon flying to LA.*
- *Time flies like an arrow.*
- *Horse flies like a sugar cube.*
- *One morning I shot an elephant in my pajamas.
How he got into my pajamas, I'll never know.*



What does NLP involve?

Syntactic Tasks

- > myspace.com/pluckerswingbar = myspace .com pluckers wing bar
- > independently \Rightarrow in + (depend + ent) + ly
- > I (Pro) ate (V) the (Det) spaghetti (N) with (Prep) meatballs (N).
- > [NP He] [VP reckons] [NP the current account deficit] [VP will narrow] [PP to] [NP only # 1.8 billion] [PP in] [NP September]
- >



Terms?

- > Segmentation
- > Morphology
- > POS tagging
- > Phrase chunking
- >
- > Syntactic parsing



What does NLP involve?

Semantic Tasks

- > Ellen has a strong interest in computational linguistics
Ellen pays a large amount of interest on her credit card.
- > agent patient source destination instrument
- > How many cities are there in the US?
answer(A, count(B, (city(B), loc(B, C),
const(C, countryid(USA))), A))

>	TEXT	HYPOTHESIS	ENTAILMENT
	<i>Eyeing the huge market potential, currently led by Google, Yahoo took over search company Overture Services Inc last year.</i>	<i>Yahoo bought Overture.</i>	TRUE

Terms?

- > Word-sense disambiguation
- > Semantic role labeling
- > Semantic parsing
- > Textual entailment



What does NLP involve?

people organizations places

- Michael Dell is the CEO of Dell Computer Corporation and lives in Austin Texas.

Information Extraction

Who was president when Barack Obama was born?

- John F. Kennedy

Question Answering



What does NLP involve?

Article: With a split decision in the final two primaries and a flurry of superdelegate endorsements, Sen. Barack Obama sealed the Democratic presidential nomination last night after a grueling and history-making campaign against Sen. Hillary Rodham Clinton that will make him the first African American to head a major-party ticket. Before a chanting and cheering audience in St. Paul, Minn., the first-term senator from Illinois savored what once seemed an unlikely outcome to the Democratic race with a nod to the marathon that was ending and to what will be another hard-fought battle, against Sen. John McCain, the presumptive Republican nominee....

Summary: Senator Barack Obama was declared the presumptive Democratic presidential nominee.

Text Summarization

Hasta la vista, bebé ⇒

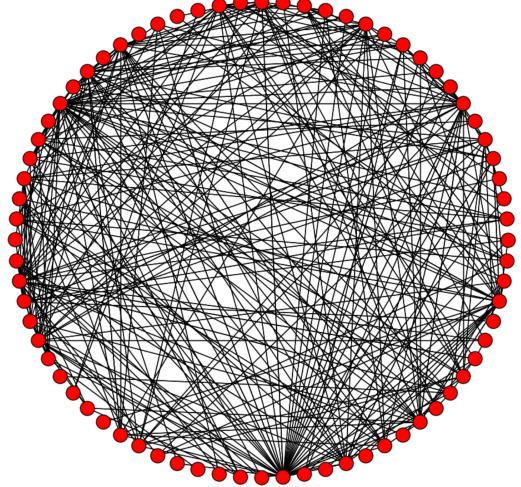
Until we see each other again, baby.

Machine Translation



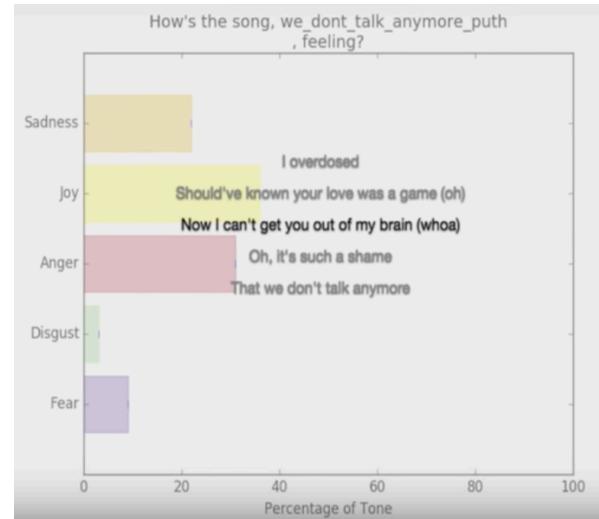
-
- > “*Consider summarization snippets on Google search results. At the moment, algorithms still use statistical models to find frequent pieces and then paste them together. With deep learning, we use complex neural networks to map text into higher-dimensional representations, and re-generate a sequence of words. All of this work has been done in the last 3-4 years.*” – Dan Jurafsky, NLP Group @ Stanford
 - > Multicore GPU’s and CPU’s
 - > “*Analysts have a more theoretical/statistical background. Therefore, [PhDs working in NLP] tend to approach problems from a mathematical standpoint. Builders work on pipelines that will handle all of the text until something is more usable to prototype.*” - Sinan Ozdemir, Kylie.ai





<http://rainier.io/mining-18000-texts-for-fun-and-no-profit>

(Judges' Favorite Hack at the 6Sense Data Hack in San Francisco)

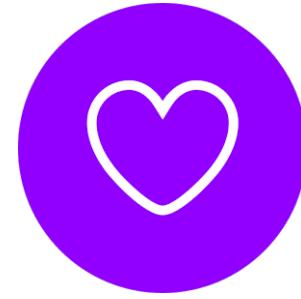
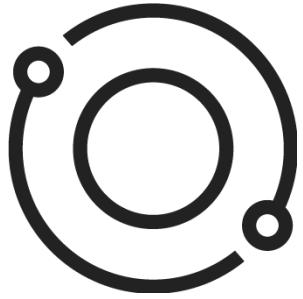


<https://github.com/theCeedo/IndigitousMusicFinder>

(Best Technology Award at Indigitous #Hack)



Some inspiration



Zebra Puzzle

From Wikipedia, the free encyclopedia

The **zebra puzzle** is a well-known [logic puzzle](#).

It is often claimed that only 2% of the population can solve the puzzle.^{[4][5]}

The following version of the puzzle appeared in *Life International* in 1962:

1. There are five houses.
2. The Englishman lives in the red house.
3. The Spaniard owns the dog.
4. Coffee is drunk in the green house.
5. The Ukrainian drinks tea.
6. The green house is immediately to the right of the ivory house.
7. The Old Gold smoker owns snails.
8. Kools are smoked in the yellow house.
9. Milk is drunk in the middle house.
10. The Norwegian lives in the first house.
11. The man who smokes Chesterfields lives in the house next to the man with the fox.
12. Kools are smoked in the house next to the house where the horse is kept.
13. The Lucky Strike smoker drinks orange juice.
14. The Japanese smokes Parliaments.
15. The Norwegian lives next to the blue house.

Now, who drinks water? Who owns the zebra?

House (1)	1	2	3	4	5
Color	Yellow (8)	Blue (15)	Red (2)	Ivory (6)	Green (4) (6) ↴
Nationality	Norwegian (15)	Ukrainian (5)	Englishman (2)	Spaniard (3)	Japanese (14)
Drink		Tea (5)	Milk (9)	Orange juice (13)	Coffee (4)
Smoke	Kools (12) (8)	Chesterfield (11)	Old Gold (7)	Lucky Strike (13)	Parliament (14)
Pet	Fox (11)	Horse (12)	Snails (7)	Dog (3)	

<https://github.com/rainiera/nlpsat> (WIP)



PowerPoint Karaoke!



PowerPoint Karaoke instructions

- > Organize into teams of 1-3 people
- > Draw for random speaking order, give random presentations to each team
- > Draw for random presentation order – give them 3 minutes to prepare – basically while the previous team is speaking.
- > Each team gets 3 minutes to speak. Timing cues at 2, 2.5 and 3 minutes
- > Q&A for each presentation

