Introduction to Natural Language Processing

DSIndy January 2018

What is NLP?

- Fundamental goal: *deep* understand of *broad* language
  - Not just string processing or keyword matching!
- End systems that we want to build:
  - Simple: spelling correction, text categorization…
  - Complex: speech recognition, machine translation, information extraction, dialog interfaces, question answering…
  - Unknown: human-level comprehension (is this just NLP?)

NLP

Natural language processing (NLP) is an area of computer science and artificial intelligence concerned with the interactions between computers and human (natural) languages, in particular how to program computers to process and analyze large amounts of natural language data.

Includes many different techniques for interpreting human language, ranging from statistical and machine learning methods to rules-based and algorithmic approaches.

History

Some Early NLP History

- 1950's:
  - Foundational work: automata, information theory, etc.
  - First speech systems
  - Machine translation (MT) hugely funded by military
    - Toy models: MT using basically word-substitution
  - Optimism!
- 1960's and 1970's: NLP Winter
  - Bar-Hillel (FAHQT) and ALPAC reports kills MT
  - Work shifts to deeper models, syntax
  - … but toy domains / grammars (SHRDLU, LUNAR)
- 1980's and 1990's: The Empirical Revolution
  - Expectations get reset
  - Corpus-based methods become central
  - Deep analysis often traded for robust and simple approximations
  - *Evaluate everything*
- 2000+: Richer Statistical Methods
  - Models increasingly merge linguistically sophisticated representations with statistical methods, confluence and clean-up
  - *Begin to get both breadth and depth*

Two Generations of NLP

- Hand-crafted Systems – Knowledge Engineering [1950s–]
  - Rules written by hand; adjusted by error analysis
  - Require experts who understand both the systems and domain
  - Iterative guess-test-tweak-repeat cycle

- Automatic, Trainable (Machine Learning) System [1985s–]
  - The tasks are modeled in a statistical way
  - More robust techniques based on rich annotations
  - Perform better than rules (Parsing 90% vs. 75% accuracy)

Early 1950s

- Machine Translation (MT): one of the earliest applications of computers
  - Major players: US and USSR
  - Russian to English and reverse
- Georgetown University, Washington system:
  - translated sample texts in 1954
  - euphoria - a lot of funding, many groups in US, USSR
  - BUT: the system could not scale up

SHRDLU

- interaction with a robot in a block world.
  - author: Terry Winograd – MIT
- the user can:
  - ask the robot to manipulate the blocks
  - ask it about the blocks configurations
  - ask it about its reasoning
  - update facts
- "understands" language in a limited domain by using syntactic parsing and semantic reasoning
  - large scale grammar of English + parser
  - procedural semantics for words and phrases

ELIZA

- the first chatterbot – a computer program that mimics human conversation
  - author: Joseph Weizenbaum –Massachusetts Institute of Technology
- simulation of a (Rogerian) therapist
  - user types in some statement or set of statements in natural language
  - ELIZA then analyzes the user's statement and generates some response
- basic technology: pattern matching

USER:	You don't argue with me.
ELIZA:	WHY DO YOU THINK I DON'T ARGUE WITH YOU

ALPAC Report

- assessed research results of groups working on MTs
- conclusions:
  - MT not possible in near future
  - funding should cease for MT!
  - basic research should be supported
- word by word translation does not work, linguistic knowledge is needed

John McArthy

1956 coined the term “Artificial Intelligence”

1958 created the LISP programming language which is still in use today

Shakey

SRI robot, Shakey, demonstrated combining locomotion, perception and problem solving. 1969

Deep Blue

1997, the first computer chess-playing system to beat a reigning world chess champion, Garry Kasparov .

DARPA award given to Stanford in 2005

driving autonomously for 131 miles along an unrehearsed desert trail

Watson

Q/A system

Overview

Two trends

An enormous amount of knowledge is now available in machine readable form as natural language text

Conversational agents are becoming an important

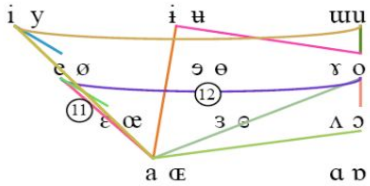
NLP, as in many areas of AI:

- We're often dealing with ill-defined problems
- We don't often come up with perfect solutions/algorithms
- We can't let either of those facts get in our way

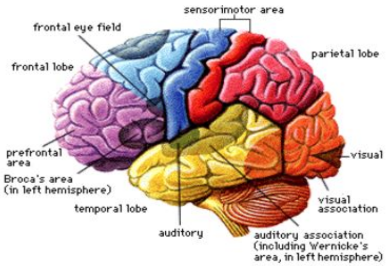
- Issue 1: Lexicon Size
  - Potentially HUGE!
  - Controlling factor: morphology
    - Store base forms (roots/stems)
      - Use morphologic process to generate / analyze
      - E.g. Dog: dog(s); sing: sings, sang, sung, singing, singer,..
- Issue 2: Lexical ambiguity
  - rock: N/V; dog: N/V;
  - "Time flies like a banana"

Related Fields

- Computational Linguistics
  - Using computational methods to learn more about how language works
  - We end up doing this and using it



- Cognitive Science
  - Figuring out how the human brain works
  - Includes the bits that do language
  - Humans: the only working NLP prototype!

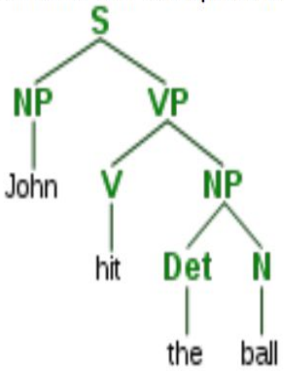


- Speech?
  - Mapping audio signals to text
  - Traditionally separate from NLP, converging?
  - Two components: acoustic models and language models
  - Language models in the domain of stat NLP

Parse trees and grammars

Parse trees

Given a grammar, a sentence can be represented as a parse tree



Grammars

Set of re-write rules, e.g.:

$$\begin{aligned} S &:= NP\ VP \\ NP &:= \text{noun}|\text{pronoun} \\ noun &:= \text{intelligence}|\text{wumpus}|\ldots \\ VP &:= \text{verb}|\text{verb}NP|\ldots \\ &\ \ldots \end{aligned}$$

- Lexicon:
  - List of legal words in a language
  - Part of speech:
    - noun, verb, adjective, determiner
- Example:
  - Noun -> cat | dog | mouse | ball | rock
  - Verb -> chase | bite | fetch | bat
  - Adjective -> black | brown | furry | striped | heavy
  - Determiner -> the | that | a | an

- What does it mean to know a language?
  - Know the words (lexicon)
    - Pronunciation, Formation, Conjugation
  - Know how the words form sentences
    - Sentence structure, Compositional meaning
  - Know how to interpret the sentence
    - Statement, question,..
  - Know how to group sentences
    - Narrative coherence, dialogue

NLP Steps

- Pick a problem (usually some disambiguation)
- Get a lot of data (usually a labeled corpus)
- Build the simplest thing that could possibly work
- Repeat:
  - See what the most common errors are
  - Figure out what information a human would use
  - Modify the system to exploit that information
    - Feature engineering
    - Representation design
    - Machine learning/statistics

Learning Methods for NLP

- **Supervised**: identify hidden units (concepts) of explicit units
  - Syntactic analysis, word sense disambiguation, name classification, relations, categorization, …
  - Trained from labeled data
- **Unsupervised**: identify relationships and properties of explicit units (terms, docs)
  - Association, topicality, similarity, clustering
  - Without labeled data
- Semi-supervised: Combinations

Concepts and Tasks

Chomsky Hierarchy of Grammar	Hierarchy of Automata
Regular Grammar	Finite State Automata
Context Free Grammar	Push Down Automata
Context Sensitive Grammar	Linear Bounded Automata
Type 0 Grammar	Turing Machine

Computationally more complex, Less Efficiency

General Framework of NLP

John runs.

Morphological and Lexical Processing

⇩

Syntactic Analysis

⇩

Semantic Analysis

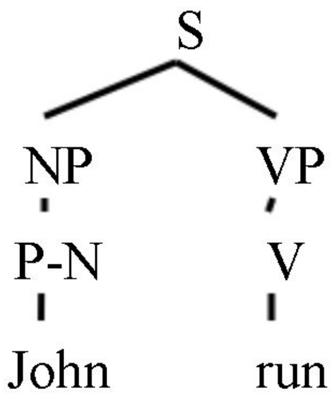
⇩

Context processing Interpretation

John run+s.

P-N V 3-pre

N plu



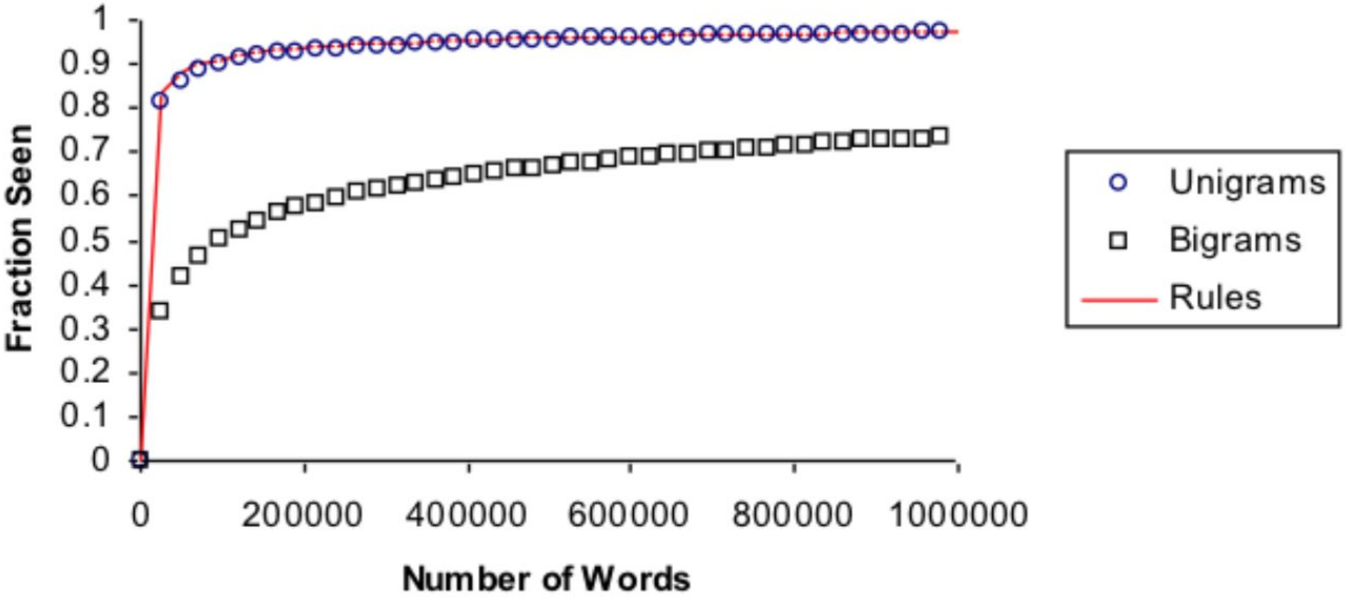
[Pred: RUN
Agent:John]

John is a student.

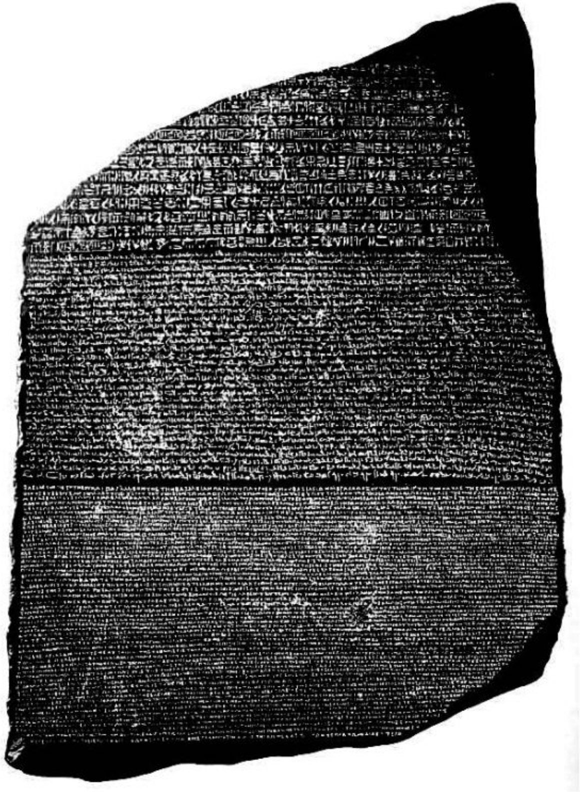
He runs.

Problem: Sparsity

- However: sparsity is always a problem
- New unigram (word), bigram (word pair), and rule rates in newswire



Corpora



A corpus is a collection of text

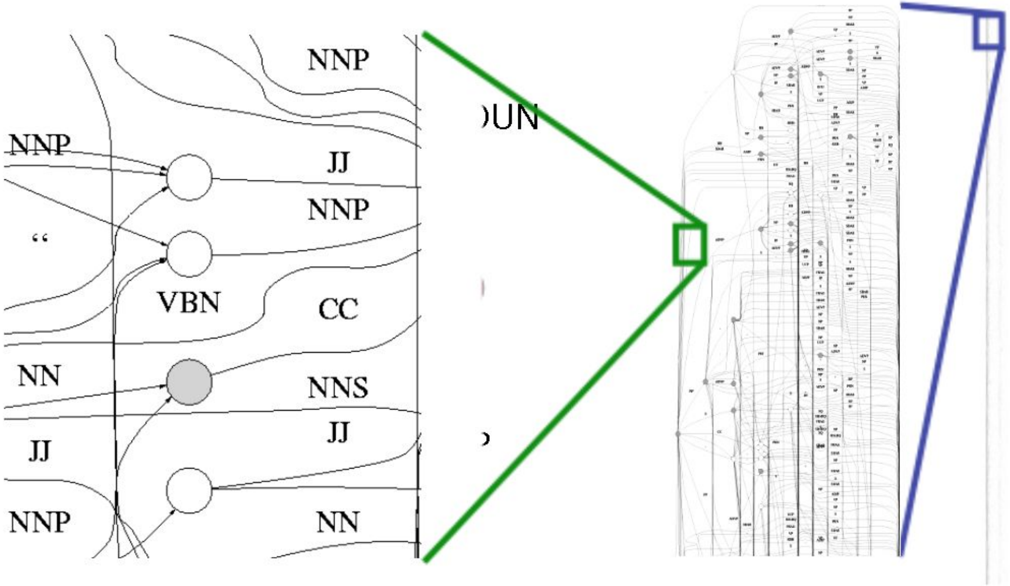
- Often annotated in some way
- Sometimes just lots of text
- Balanced vs. uniform corpora

Examples

- Newswire collections: 500M+ words
- Brown corpus: 1M words of tagged "balanced" text
- Penn Treebank: 1M words of parsed WSJ
- Canadian Hansards: 10M+ words of aligned French / English sentences
- The Web: billions of words of who knows what

Problem: Scale

- People *did* know that language was ambiguous!
- …but they hoped that all interpretations would be “good” ones (or ruled out pragmatically)
- …they didn’t realize how bad it would be

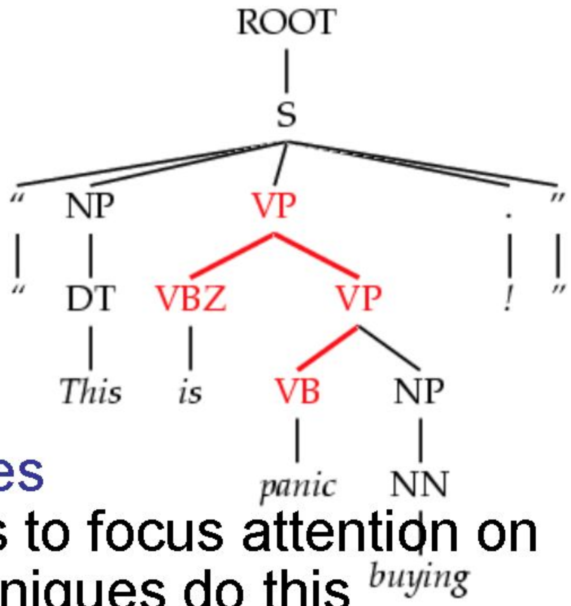


Dark Ambiguities

Dark ambiguities: most structurally permitted analyses are so bad that you can't get your mind to produce them

This analysis corresponds to the correct parse of

"This will panic buyers ! "



Unknown words and new usages

Solution: We need mechanisms to focus attention on the best ones, probabilistic techniques do this

Semantic Ambiguity

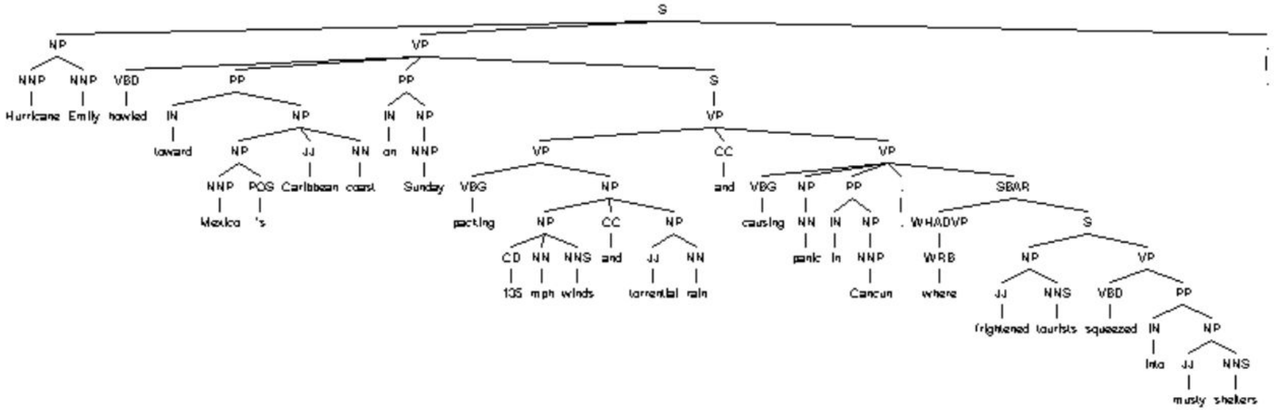
- NLP is much more than syntax!
- Even correct tree structured syntactic analyses don't fully nail down the meaning

Every morning someone's alarm clock wakes me up

John's boss said he was doing better

- In general, every level of linguistic structure comes with its own ambiguities…

Syntactic Analysis



Hurricane Emily howled toward Mexico 's Caribbean coast on Sunday packing 135 mph winds and torrential rain and causing panic in Cancun , where frightened tourists squeezed into musty shelters .

- SOTA: ~90% accurate for many languages when given many training examples, some progress in analyzing languages given few or no examples

Problem: Ambiguities

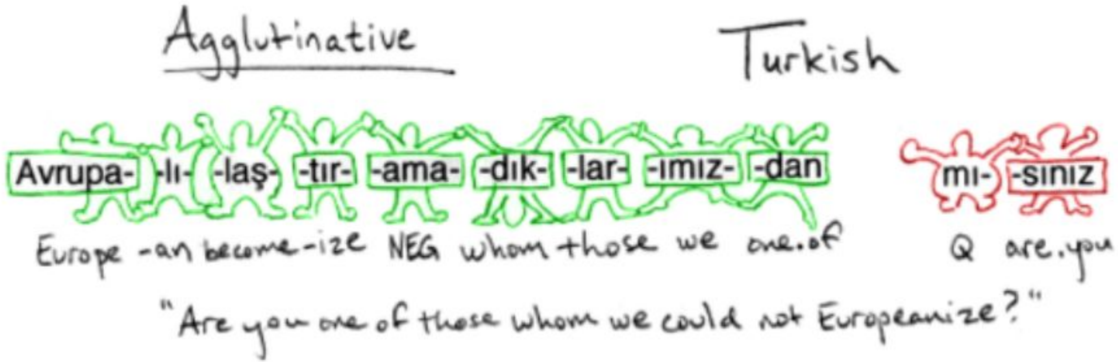
- Headlines:
  - Enraged Cow Injures Farmer with Ax
  - Ban on Nude Dancing on Governor's Desk
  - Teacher Strikes Idle Kids
  - Hospitals Are Sued by 7 Foot Doctors
  - Iraqi Head Seeks Arms
  - Stolen Painting Found by Tree
  - Kids Make Nutritious Snacks
  - Local HS Dropouts Cut in Half
- Why are these funny?

Coreference

- But the little prince could not restrain admiration:
- "Oh! How beautiful you are!"
- "Am I not?" the flower responded, sweetly. "And I was born at the same moment as the sun . . ."
- The little prince could guess easily enough that she was not any too modest--but how moving--and exciting--she was!
- "I think it is time for breakfast," she added an instant later. "If you would have the kindness to think of my needs--"
- And the little prince, completely abashed, went to look for a sprinkling-can of fresh water. So, he tended the flower.

Morphology

- The study of internal structures of words and how they can be modified
- Parsing complex words into their components



Phonetics and phonology

- The study of linguistic sounds and their relations to words

Das Funkalphabet - German Phonetic Spelling Code compared to the international ICAO/NATO code **Listen to AUDIO for this chart!** (below)		
Germany*	**Phonetic Guide**	**ICAO/NATO****
A wie **Anton**	AHN-tone	**Alfa/Alpha**
Ä wie **Ärger**	AIR-gehr	(1)
B wie **Berta**	BARE-tuh	**Bravo**
C wie **Cäsar**	SAY-zar	**Charlie**
Ch wie **Charlotte**	shar-LOT-tuh	(1)
D wie **Dora**	DORE-uh	**Delta**
E wie **Emil**	ay-MEAL	**Echo**
F wie **Friedrich**	FREED-reech	**Foxtrot**
G wie **Gustav**	GOOS-tahf	**Golf**
H wie **Heinrich**	HINE-reech	**Hotel**
I wie **Ida**	EED-uh	**India/Indigo**
J wie **Julius**	YUL-ee-oos	**Juliet**
K wie **Kaufmann**	KOWF-mann	**Kilo**
L wie **Ludwig**	LOOD-vig	**Lima**
AUDIO 1 > Listen to mp3 for A-L		
M wie **Martha**	MAR-tuh	**Mike**
N wie **Nordpol**	NORT-pole	**November**
O wie **Otto**	AHT-toe	**Oscar**
Ö wie **Ökonom** (2)	UEH-ko-nome	(1)
P wie **Paula**	POW-luh	**Papa**
Q wie **Quelle**	KVEL-uh	**Quebec**
R wie **Richard**	REE-shart	**Romeo**
S wie **Siegfried** (3)	SEEG-freed	**Sierra**
Sch wie **Schule**	SHOO-luh	(1)
ß (**Eszett**)	ES-TSET	(1)
T wie **Theodor**	TAY-oh-dore	**Tango**
U wie **Ulrich**	OOL-reech	**Uniform**
Ü wie **Übermut**	UEH-ber-moot	(1)
V wie **Viktor**	VICK-tor	**Victor**
W wie **Wilhelm**	VIL-helm	**Whiskey**
X wie **Xanthippe**	KSAN-tipp-uh	**X-Ray**
Y wie **Ypsilon**	IPP-see-lohn	**Yankee**
Z wie **Zeppelin**	TSEP-puh-leen	**Zulu**

http://german.about.com/library/blfunkabc.htm

Word sense disambiguation

- Figuring out the exact meaning of a word or entity

oun 1. **tie** - neckwear consisting of a long narrow piece of material worn (mostly by men) under a collar and tied in knot at the front; "he stood in front of the mirror tightening his necktie"; "he wore a vest and tie"

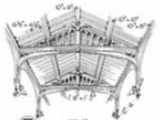
- necktie
- bola, bola tie, bolo, bolo tie - a cord fastened around the neck with an ornamental clasp and worn as a necktie
- bow tie, bow-tie, bowtie - a man's tie that ties in a bow
- four-in-hand - a long necktie that is tied in a slipknot with one end hanging in front of the other
- neckwear - articles of clothing worn about the neck
- old school tie - necktie indicating the school the wearer attended
- string tie - a very narrow necktie usually tied in a bow
- Windsor tie - a wide necktie worn in a loose bow

2. **tie** - a social or business relationship; "a valuable financial affiliation"; "he was sorry he had to sever his ties with other members of the team"; "many close associations with England"

- affiliation, tie-up, association
- relationship - a state involving mutual dealings between people or parties or countries

3. **tie** - equality of score in a contest
- equivalence, par, equality, equation - a state of being essentially equal or equivalent; equally balanced; "on a par with the best"
- deuce - a tie in tennis or table tennis that requires winning two successive points to win the game

4. **tie** - a horizontal beam used to prevent two other structural members from spreading apart or separating; "he nailed the rafters together with a tie beam"
- tie beam
- beam - long thick piece of wood or metal or concrete, etc., used in construction

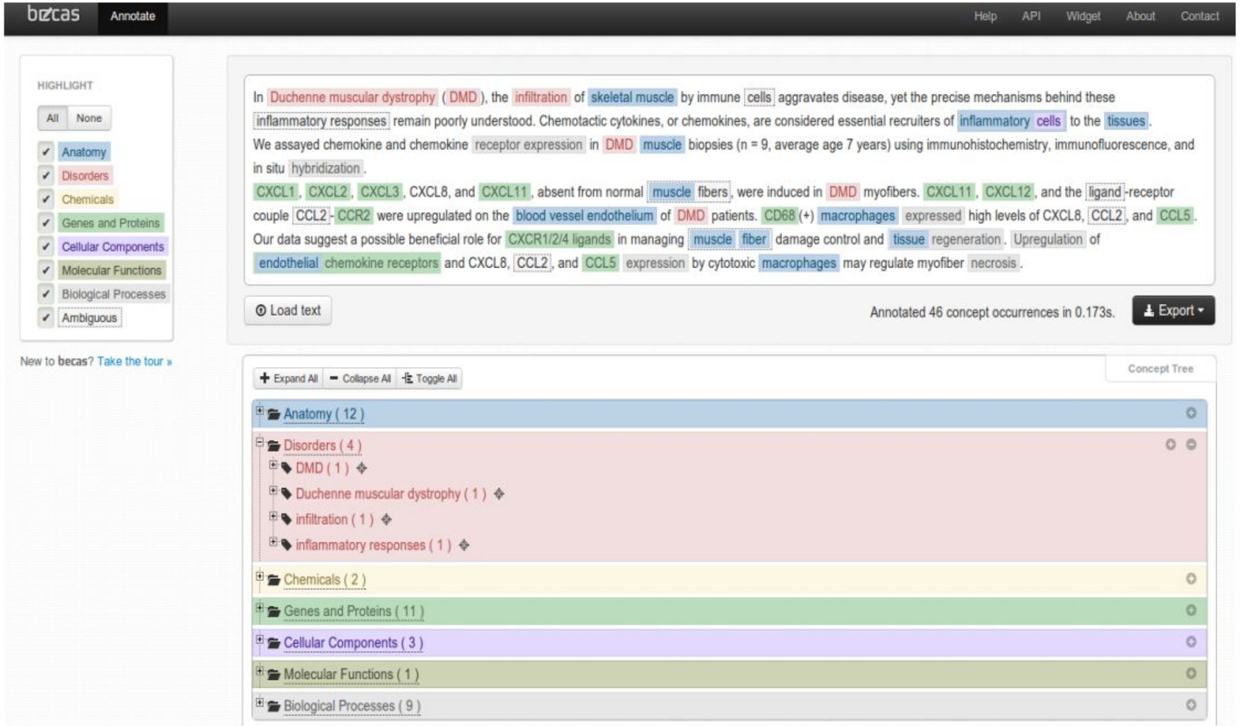


Temporal Information Extraction

- *In **1975**, after being fired from Columbia amid allegations that he used company funds to pay for his son's bar mitzvah, **Davis** founded **Arista***
  - Is ‘1975’ related to the employee_of relation between Davis and Arista?
  - If so, does it indicate START, END, HOLDS… ?
- Each classification instance represents a temporal expression in the context of the entity and slot value.
- We consider the following classes
  - START *Rob joined Microsoft in 1999.*
  - END *Rob left Microsoft in 1999.*
  - HOLDS *In 1999 Rob was still working for Microsoft.*
  - RANGE *Rob has worked for Microsoft for the last ten years.*
  - NONE *Last Sunday Rob’ s friend joined Microsoft.*

Named-entity recognition

- Identifying pre-defined entity types in a sentence



Part-of-speech tagging

- Assigning a syntatic tag to each word in a sentence

Stanford Parser

Please enter a sentence to be parsed:

```
Surgical resection specimens of 85 invasive ductal
carcinomas of 85 women who had undergone 3D
ultrasound were included.
```

Language: English ▾ Sample Sentence Parse

Your query

Surgical resection specimens of 85 invasive ductal carcinomas of 85 women who had undergone 3D ultrasound were included.

Tagging

```
Surgical/NNP resection/NN specimens/NNS of/IN 85/CD invasive/JJ
ductal/JJ carcinomas/NNS of/IN 85/CD women/NNS who/WP had/VBD
undergone/VBN 3D/CD ultrasound/NN were/VBD included/VBN ./.
```

Sentiment Analysis

- Identifying sentiments and opinions stated in a text

Question answering

- Answering questions with a short answer

===> what countries speak Spanish

The language Spanish is spoken in Argentina, Aruba, Belize, Bolivia, Brazil, Canada, Cayman Islands, Chile, Colombia, Costa Rica, Cuba, Curacao, Dominican Republic, Ecuador, El Salvador, Equatorial Guinea, Falkland Islands (Islas Malvinas), Gibraltar, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Puerto Rico, Saint Martin, Sint Maarten, Spain, Switzerland, Trinidad and Tobago, United States, Uruguay, Venezuela, and Virgin Islands.

The language Castilian Spanish is spoken in Spain.

Text Categorization

- Assigning one (or more) pre-defined category to a text

PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed
Advanced

Display Settings: Abstract

Send to:

Nature. 2014 Mar 20;507(7492):323-8. doi: 10.1038/nature13145. Epub 2014 Mar 12.

Coupling of angiogenesis and osteogenesis by a specific vessel subtype in bone.

Kusumbe AP[1], Ramasamy SK[1], Adams RH[2].

Author information

Abstract

The mammalian skeletal system harbours a hierarchical system of mesenchymal stem cells, osteoprogenitors and osteoblasts sustaining lifelong bone formation. Osteogenesis is indispensable for the homeostatic renewal of bone as well as regenerative fracture healing, but these processes frequently decline in ageing organisms, leading to loss of bone mass and increased fracture incidence. Evidence indicates that the growth of blood vessels in bone and osteogenesis are coupled, but relatively little is known about the underlying cellular and molecular mechanisms. Here we identify a new capillary subtype in the murine skeletal system with distinct morphological, molecular and functional properties. These vessels are found in specific locations, mediate growth of the bone vasculature, generate distinct metabolic and molecular microenvironments, maintain perivascular osteoprogenitors and couple angiogenesis to osteogenesis. The abundance of these vessels and associated osteoprogenitors was strongly reduced in bone from aged animals, and pharmacological reversal of this decline allowed the restoration of bone mass.

Comment in

Bone biology: Vessels of rejuvenation. [Nature. 2014]

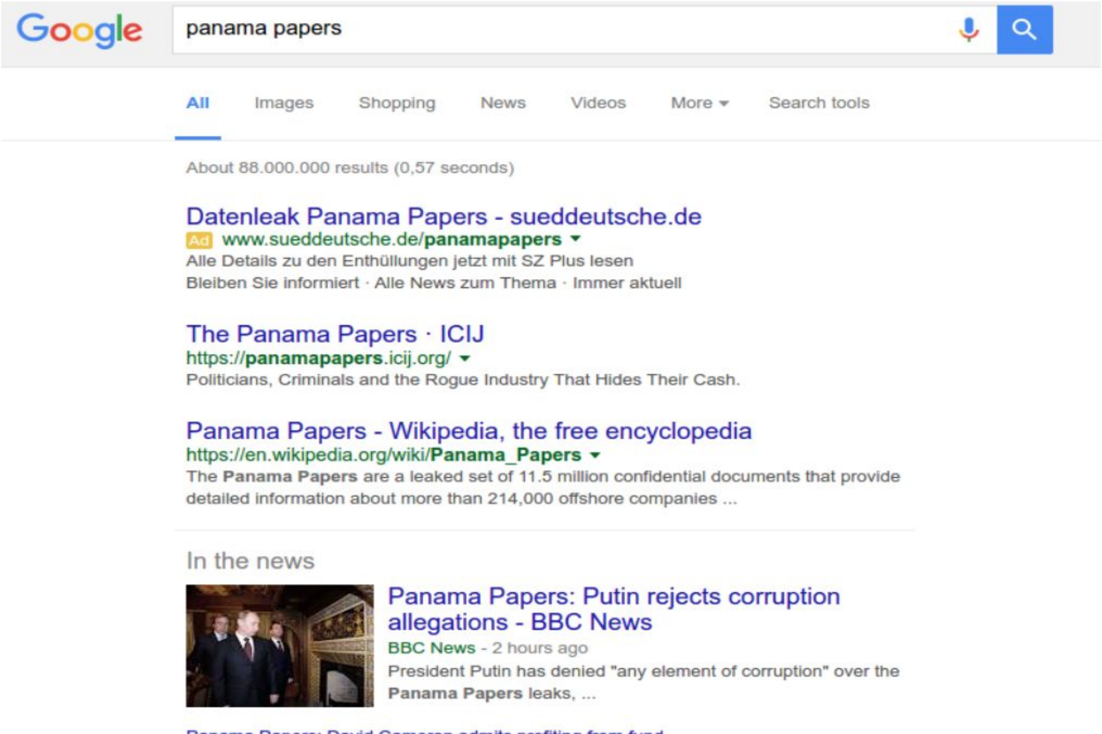
PMID: 24646994 [PubMed - indexed for MEDLINE]

MeSH Terms

Aging/metabolism
Aging/pathology
Animals
Blood Vessels/anatomy & histology
Blood Vessels/cytology
Blood Vessels/growth & development
Blood Vessels/physiology*
Bone and Bones/blood supply*
Bone and Bones/cytology
Endothelial Cells/metabolism
Hypoxia-Inducible Factor 1, alpha Subunit/metabolism
Male
Mice
Mice, Inbred C57BL
Neovascularization, Physiologic/physiology*
Osteoblasts/cytology
Osteoblasts/metabolism
Osteogenesis/physiology*
Oxygen/metabolism
Stem Cells/cytology
Stem Cells/metabolism

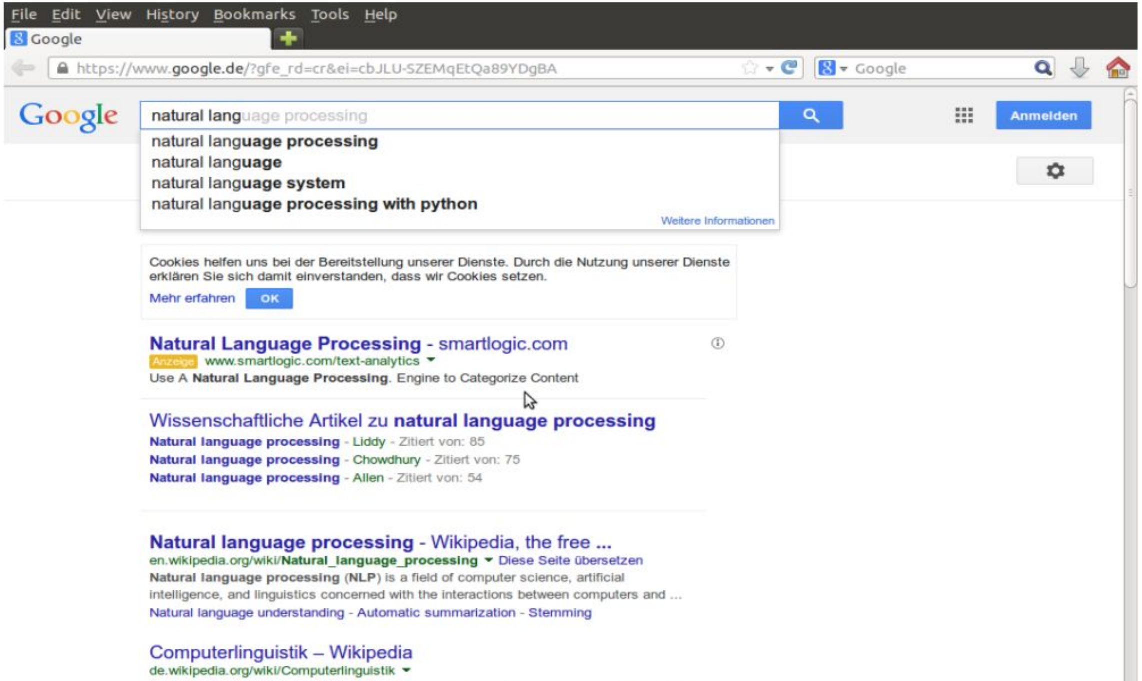
Information Retrieval

- Finding relevant information to the user's query



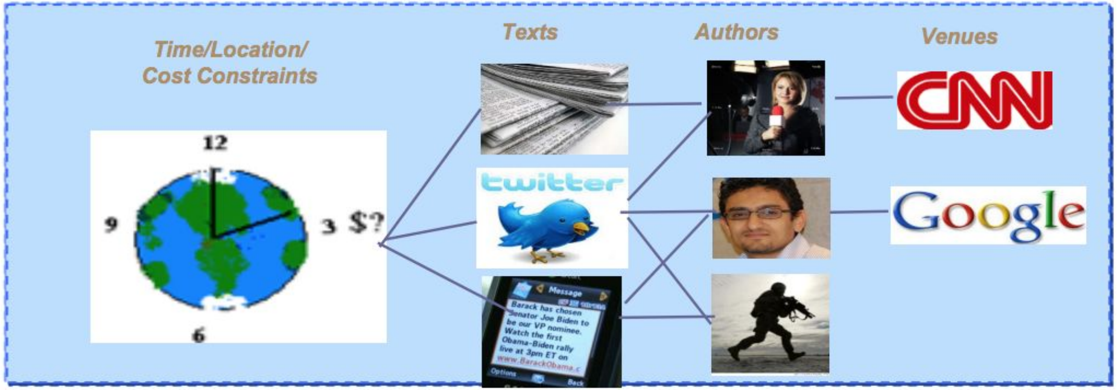
Word Prediction

- Predicting the next word that is highly probable to be typed by the user



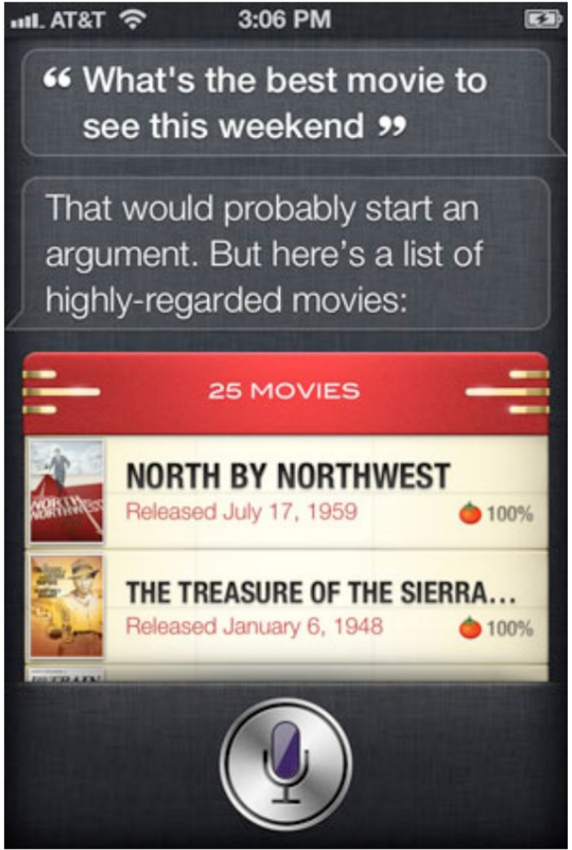
NLP for Big Data

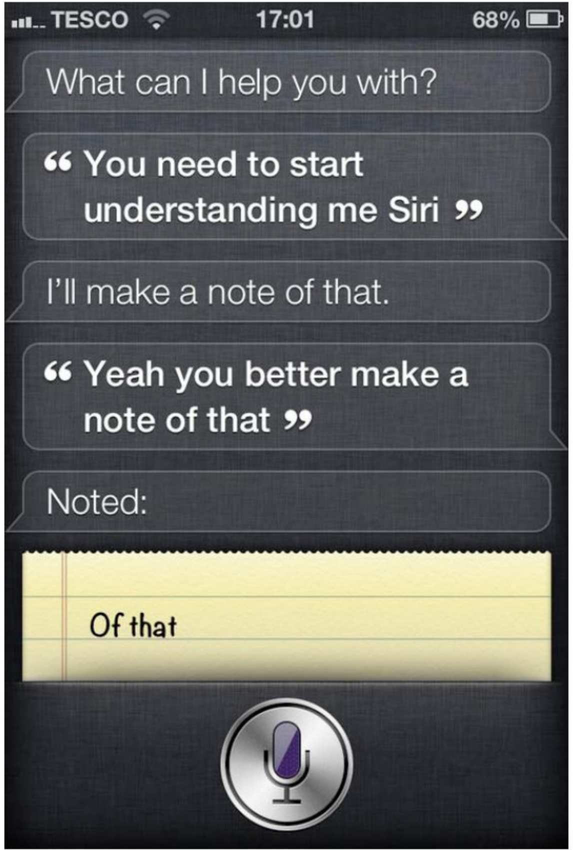
- Huge Size
  - Google processes 20 PB a day (2008)
  - Wayback Machine has 3 PB + 100 TB/month (3/2009)
  - Facebook has 2.5 PB of user data + 15 TB/day (4/2009)
  - eBay has 6.5 PB of user data + 50 TB/day (5/2009)
- 80% data is unstructured (IBM, 2010)
- More importantly, Heterogeneous



Current Work

Mobile devices can now answer (some or our) questions and execute commands...





Jeopardy! World Champion

US Cities: Its largest airport is named for a World War II hero; its second largest, for a World War II battle.

Semantic tasks (how would you reduce these to prediction problems?)

- Sentiment analysis
- Summarization
- Information extraction, slot-filling
- Discourse analysis
- Textual entailment

Speech:

- Better language modeling (predict next word) – syntax, semantics
- Better models of acoustics, pronunciation
  - fewer speaker-specific parameters
    - to enable rapid adaptation to new speakers
  - more robust recognition
    - emotional speech, informal conversation, meetings
    - juvenile/elderly voices, bad audio, background noise
  - Some techniques to solve these:
    - non-local features
    - physiologically informed models
    - dimensionality reduction

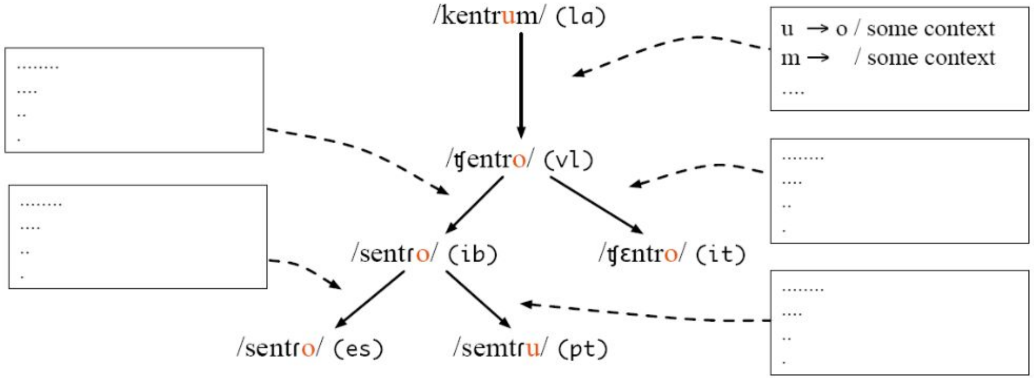
Machine translation:

- Best-funded area of NLP, right now
- Models and algorithms
- How to incorporate syntactic structure?
- "Low-resource" and morphologically complex languages?

Timeline of Deep Learning

300 BC	Aristotle	introduced Associationism, started the history of human's attempt to understand brain.
1873	Alexander Bain	introduced Neural Groupings as the earliest models of neural network, inspired Hebbian Learning Rule.
1943	McCulloch & Pitts	introduced MCP Model, which is considered as the ancestor of Artificial Neural Model.
1949	Donald Hebb	considered as the father of neural networks, introduced Hebbian Learning Rule, which lays the foundation of modern neural network.
1958	Frank Rosenblatt	introduced the first perceptron, which highly resembles modern perceptron.
1974	Paul Werbos	introduced Backpropagation
1980	Teuvo Kohonen	introduced Self Organizing Map
	Kunihiko Fukushima	introduced Neocogitron, which inspired Convolutional Neural Network
1982	John Hopfield	introduced Hopfield Network
1985	Hilton & Sejnowski	introduced Boltzmann Machine
1986	Paul Smolensky	introduced Harmonium, which is later known as Restricted Boltzmann Machine
	Michael I. Jordan	defined and introduced Recurrent Neural Network
1990	Yann LeCun	introduced LeNet, showed the possibility of deep neural networks in practice
1997	Schuster & Paliwal	introduced Bidirectional Recurrent Neural Network
	Hochreiter & Schmidhuber	introduced LSTM, solved the problem of vanishing gradient in recurrent neural networks
2006	Geoffrey Hinton	introduced Deep Belief Networks, also introduced layer-wise pretraining technique, opened current deep learning era.
2009	Salakhutdinov & Hinton	introduced Deep Boltzmann Machines
2012	Geoffrey Hinton	introduced Dropout, an efficient way of training neural networks

Etc: Historical Change



Gloss	Latin	Italian	Spanish	Portuguese
Word/verb	verbum	verbo	verbo	verbu
Center	centrum	centro	centro	centro

- Change in form over time, reconstruct ancient forms, phylogenies
- … just an example of the many other kinds of models we can build

Machine Translation

"Il est impossible aux journalistes de rentrer dans les régions tibétaines"

Bruno Philip, correspondant du "Monde" en Chine, estime que les journalistes de l'AFP qui ont été expulsés de la province tibétaine du Qinghai "n'étaient pas dans l'illégalité".

Les faits Le dalaï-lama dénonce l'"enfer" imposé au Tibet depuis sa fuite, en 1959

Vidéo Anniversaire de la rébellion

"It is impossible for journalists to enter Tibetan areas"

Philip Bruno, correspondent for "World" in China, said that journalists of the AFP who have been deported from the Tibetan province of Qinghai "were not illegal."

Facts The Dalai Lama denounces the "hell" imposed since he fled Tibet in 1959

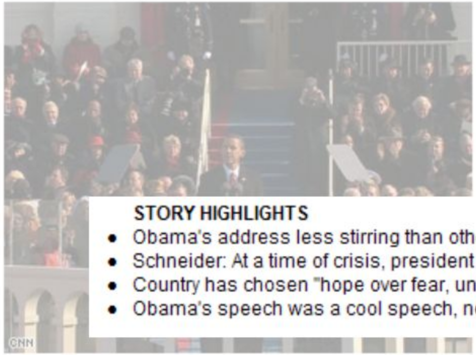
Video Anniversary of the Tibetan rebellion: China on guard

- Translate text from one language to another
- Recombines fragments of example translations
- Challenges:
  - What fragments? [learning to translate]
  - How to make efficient? [fast translation search]
  - Fluency (next class) vs fidelity (later)

Summarization

- Condensing documents
  - Single or multiple
  - Extractive or synthetic
  - Aggregative or representative
  - Even just shortening sentences
- Very context-dependent!
- An example of analysis with generation

WASHINGTON (CNN) -- President Obama's inaugural address was cooler, more measured and reassuring than that of other presidents making it, perhaps, the right speech for the times.



President Obama renewed his call for a massive plan to stimulate economic growth.

more photos »

STORY HIGHLIGHTS
- Obama's address less stirring than others but more candid, analyst says
- Schneider: At a time of crisis, president must be reassuring
- Country has chosen "hope over fear, unity of purpose over ... discord," Obama said
- Obama's speech was a cool speech, not a hot one, Schneider says

Some inaugural addresses are known for their soaring, inspirational language. Like John F. Kennedy's in 1961: "Ask not what your country can do for you. Ask what you can do for your country."

Obama's address was less stirring, perhaps, but it was also more candid and down-to-earth.

"Starting today," the new president said, "we must ... begin

... said in his first inaugural in 1933, "The only thing we have to fear is fear itself." Or Bill Clinton, who took office during the economic crisis of the early 1990s. "There is nothing wrong with America that cannot be fixed by what is right with America," Clinton declared at his first inaugural.

Obama, too, offered reassurance.

"We gather because we have chosen hope over fear, unity of purpose over conflict and discord," Obama said.

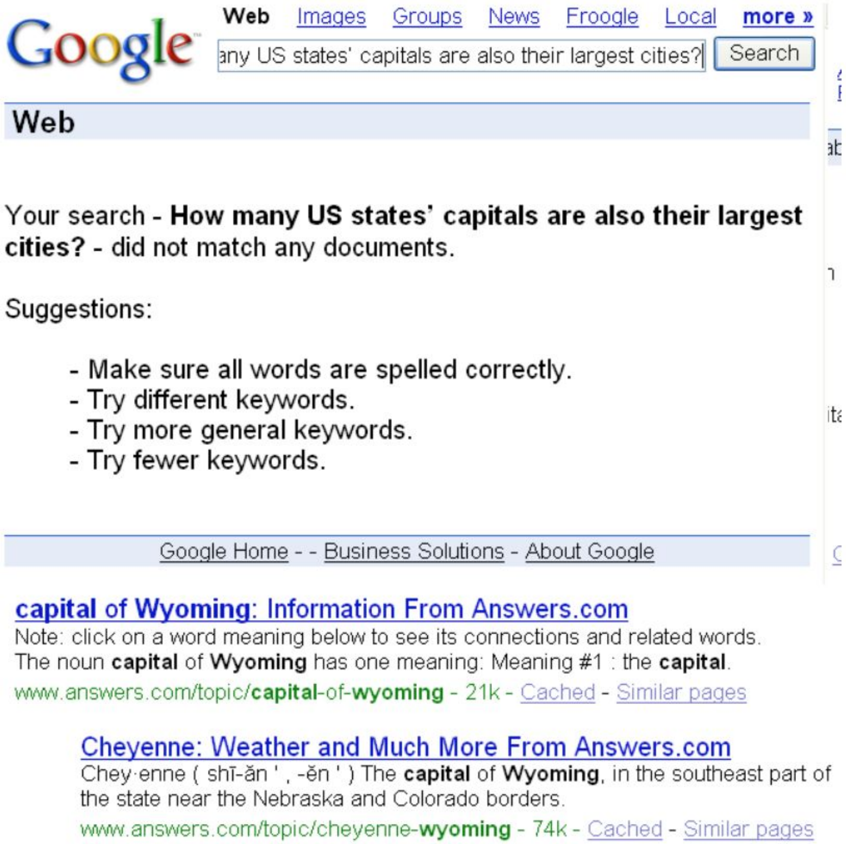
Obama's call to unity after decades of political division echoed Abraham Lincoln's first inaugural address in 1861. Even though he delivered it at the onset of a terrible civil war, Lincoln's speech was not a call to battle. It was a call to look beyond the war, toward reconciliation based on what he called "the better angels of our nature."

Some presidents used their inaugural address to set out a bold agenda.

Question Answering

- Question Answering:
  - More than search
  - Ask general comprehension questions of a document collection
  - Can be really easy: “What’s the capital of Wyoming?”
  - Can be harder: “How many US states’ capitals are also their largest cities?”
  - Can be open ended: “What are the main issues in the global warming debate?”

- SOTA: Can do factoids, even when text isn’t a perfect match



Information Extraction

- Unstructured text to database entries

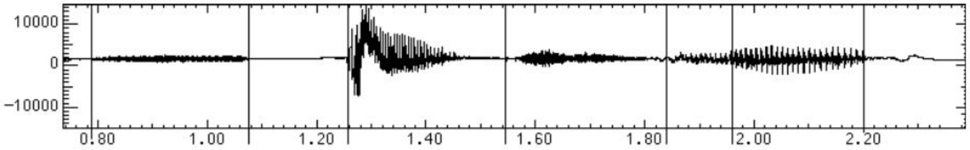
New York Times Co. named Russell T. Lewis, 45, president and general manager of its flagship New York Times newspaper, responsible for all business-side activities. He was executive vice president and deputy general manager. He succeeds Lance R. Primis, who in September was named president and chief operating officer of the parent.

Person	Company	Post	State
Russell T. Lewis	New York Times newspaper	president and general manager	start
Russell T. Lewis	New York Times newspaper	executive vice president	end
Lance R. Primis	New York Times Co.	president and CEO	start

- SOTA: perhaps 80% accuracy for multi-sentence temples, 90%+ for single easy fields
- But remember: information is redundant!

Speech Systems

- Automatic Speech Recognition (ASR)
  - Audio in, text out
  - SOTA: 0.3% error for digit strings, 5% dictation, 50%+ TV



"Speech Lab"

- Text to Speech (TTS)
  - Text in, audio out
  - SOTA: totally intelligible (if sometimes unnatural)

From symbolic to neural computation

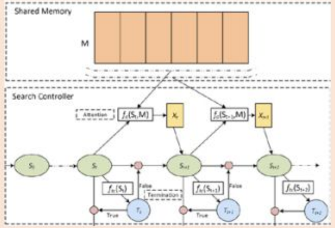
Input: Q

Question: Symbolic → Neural
by embedding models / encoder

Symbolic Space
- UI: human readable I/O
- Leverage traditional symbolic approaches as pre/post processing
  - Keyword matching
  - Ontology based models
  - e.g., doc/passage/entity search/ranking

Inference: Question + KG → Answer
by RNN + Attention

Neural Space



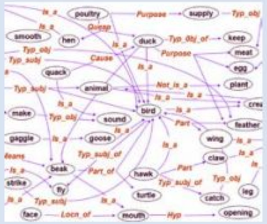
Error(A, A*)

Output: A

Answer: Neural → Symbolic
by generative models / decoder

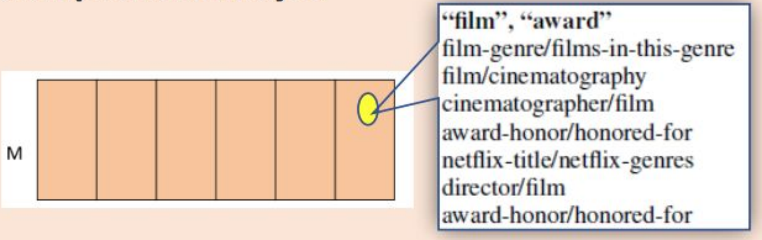
Symbolic Space

- **Knowledge Representation**
  - *Explicitly* store a BIG but incomplete knowledge graph (KG)
  - Words, relations, templates
  - High-dim, discrete, sparse vectors
- **Inference**
  - Slow on a big KG
  - Keyword/template matching is sensitive to paraphrase alternations
- **Human comprehensible but not computationally efficient**

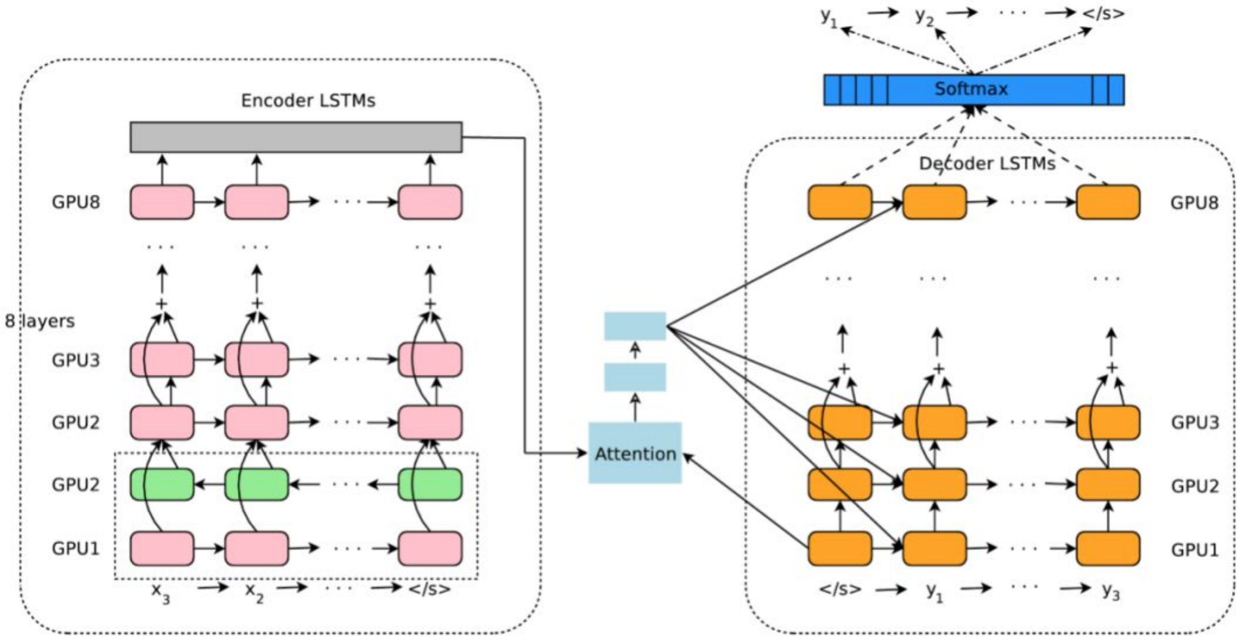


Neural Space

- **Knowledge Representation**
  - *Implicitly* store entities and structure of KG in a *compact* way that is more generalizable
  - Semantic concepts/classes
  - Low-dim, cont., dense vectors shaped by KG
- **Inference**
  - Fast on compact memory
  - Semantic matching is robust to paraphrase alternations
- **Computationally efficient but not human comprehensible *yet***



Google's NTM system



- Deep RNNs
- Residual connections
- Bi-directional encoder for first layer
- The use of sub-word units
- Model parallelism

Gated Recurrent Unit (GRU)

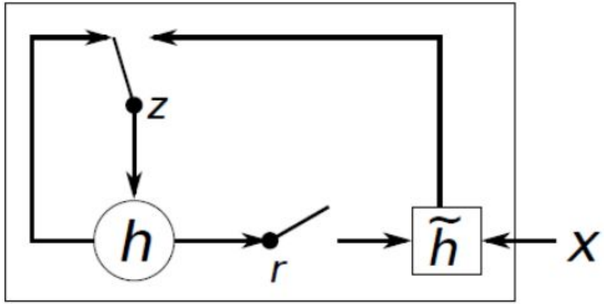


Figure 2: An illustration of the proposed hidden activation function. The update gate z selects whether the hidden state is to be updated with a new hidden state $\tilde{h}$. The reset gate r decides whether the previous hidden state is ignored. See

$$r_j = \sigma\left(\left[\mathbf{W}_r\mathbf{x}\right]_j + \left[\mathbf{U}_r\mathbf{h}_{\langle t-1\rangle}\right]_j\right)$$

$$z_j = \sigma\left(\left[\mathbf{W}_z\mathbf{x}\right]_j + \left[\mathbf{U}_z\mathbf{h}_{\langle t-1\rangle}\right]_j\right)$$

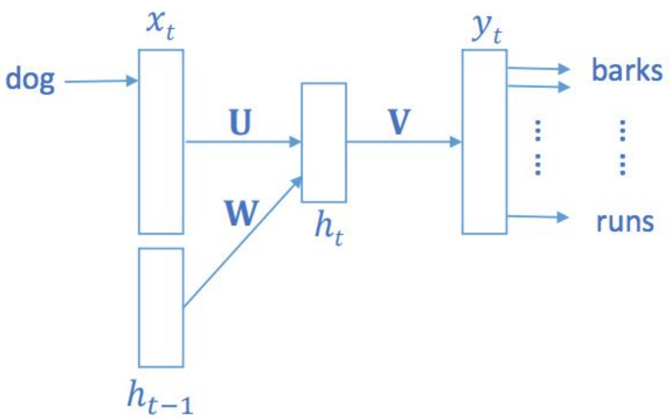
$$\tilde{h}_j^{\langle t\rangle} = \phi\left(\left[\mathbf{W}\mathbf{x}\right]_j + \left[\mathbf{U}\left(\mathbf{r}\odot\mathbf{h}_{\langle t-1\rangle}\right)\right]_j\right)$$

$$h_j^{\langle t\rangle} = z_j h_j^{\langle t-1\rangle} + (1 - z_j)\tilde{h}_j^{\langle t\rangle}$$

Distributed representation of words

- A lot of popular methods for creating word vectors!
  - Vector Space Model [Salton & McGill 83]
  - Latent Semantic Analysis [Deerwester+ 90]
  - Brown Clustering [Brown+ 92]
  - Latent Dirichlet Allocation [Blei+ 03]
  - Deep Neural Networks [Collobert & Weston 08]
  - DSSM [Huang+ 13]
  - Word2Vec [Mikolov+ 13]
  - GloVe [Pennington+ 14]
- Encode term co-occurrence information
- Measure semantic similarity well

Recurrent Neural Network (RNN) for Language Modeling



x_t: input one-hot vector at time step t
h_t: encodes the history of all words up to time step t
y_t: distribution of output words at time step t

$$z_t = \mathbf{U}x_t + \mathbf{W}h_{t-1}$$
$$h_t = \sigma(z_t)$$
$$y_t = g(\mathbf{V}h_t)$$

where

$$\sigma(z) = \frac{1}{1+\exp(-z)}, \quad g(z_m) = \frac{\exp(z_m)}{\sum_k \exp(z_k)}$$

$g(.)$ is called the *softmax* function

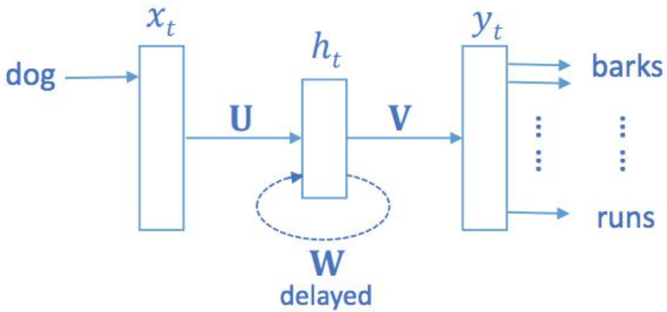
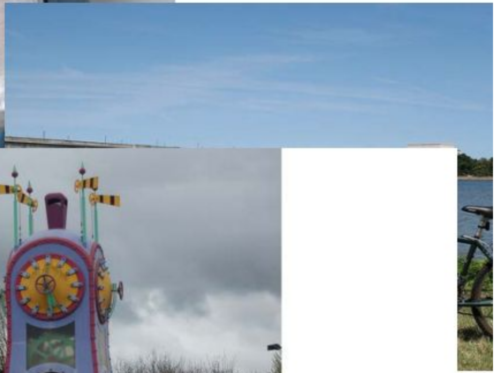


Table 1: *Performance of models on WSJ DEV set when increasing size of training data.*

Model	# words	PPL	WER
KN5 LM	200K	336	16.4
KN5 LM + RNN 90/2	200K	271	15.4
KN5 LM	1M	287	15.1
KN5 LM + RNN 90/2	1M	225	14.0
KN5 LM	6.4M	221	13.5
KN5 LM + RNN 250/5	6.4M	156	11.7

Machine-generated (but turker prefered)	a group of motorcycles parked next to a motorcycle
Human-annotated (but turker not prefered)	two girls wearing are wearing short skirts and one of them sits on a motorcycle while the other stands nearby

Machine-generated (but turker prefered)	a wom food
Human-annotated (but turker not prefered)	womar kitcher



ne

nex

Machine-generated (but turker prefered)	a clock tower in the middle of the street
Human-annotated (but turker not prefered)	a statue with a clock on it near a parking lot

DSSM: Compute Similarity in Semantic Space

Relevance measured by cosine similarity

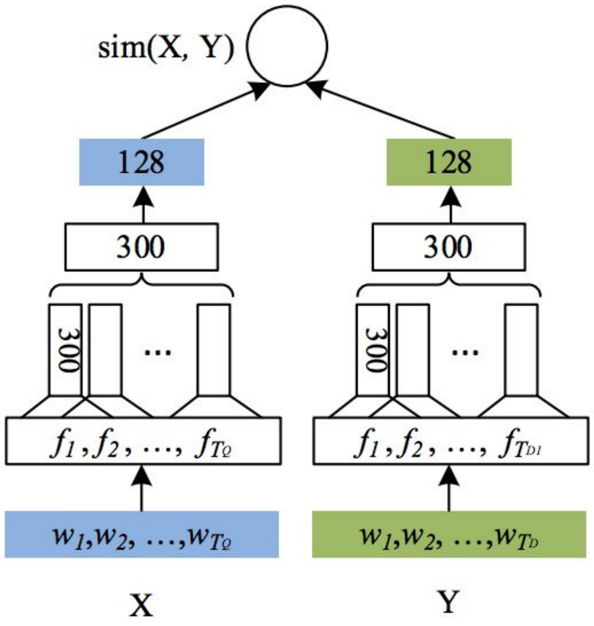
Semantic layer h

Max pooling layer v

Convolutional layer c_t

Word hashing layer f_t

Word sequence x_t



Learning: maximize the similarity between X (source) and Y (target)

Representation: use DNN to extract abstract semantic representations

Convolutional and Max-pooling layer: identify key words/concepts in X and Y

Word hashing: use sub-word unit (e.g., letter n-gram) as raw input to handle very large vocabulary

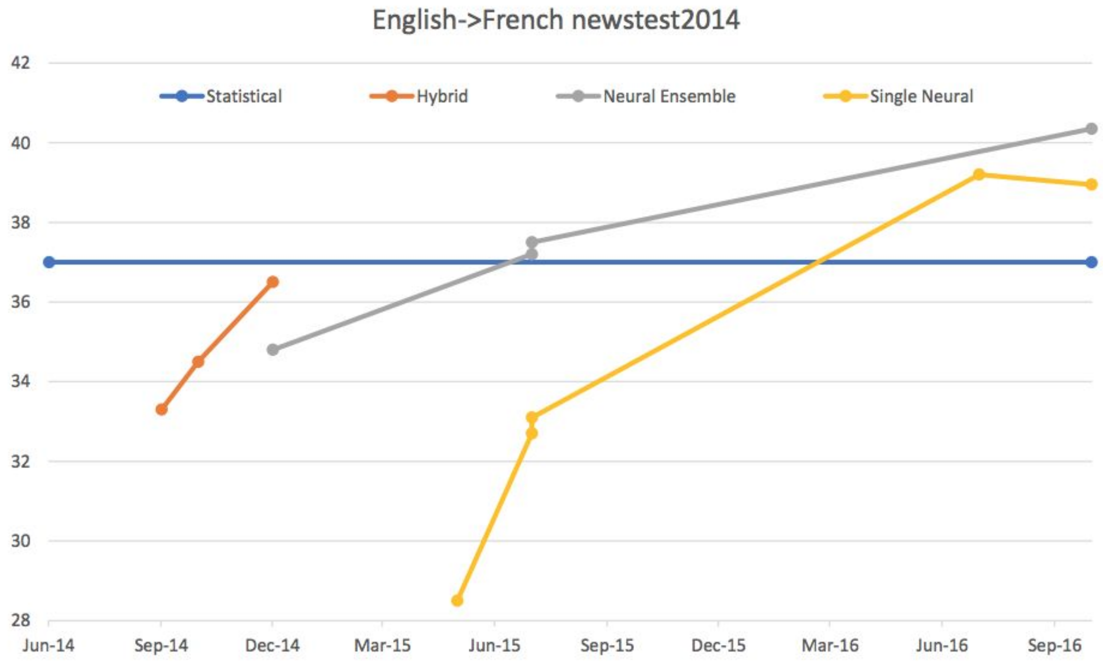
- Classification task – label x by y
  - MLP/CNN/RNN as feature generator
- Ranking task – compute the semantic similarity btw x and y
  - Siamese neural network [Bromley et al. 1993]
  - Deep Semantic Similarity Model (DSSM)
- (Text) Generation task – generate y from x
  - Seq2Seq (RNN/LSTM)
  - Memory Network
- Question answering task
  - Neural machine reading models
- Task-completion dialogue
  - Deep reinforcement learning for dialogue agents

Recent progress on machine translation (MT): BLEU score of state of the art systems

Statistical MT state of the art is highly engineered and has made little progress in over two years

Chart shows progress in three classes of neural systems

- Hybrid: Add neural models to existing statistical MT system
- Single: Single pure-neural system (to be discussed in Part 3)
- Ensemble: Large ensemble of pure-neural systems



State of the art results on NLP component tasks

Task	Test set	Metric	Best non-neural	Best neural	Source
POS tagging	PTB section 23	F1	97.17	97.78	Andor+ 16
Syntactic Parsing	PTB section 23	F1	90.1	93.3	Dyer+ 16
Dependency parsing	PTB section 23	F1	93.22	94.61	Andor+ 16
CCG parsing	CCGBank test	F1	85.2	88.7	Lee+ 16
Inference (NLI)	Stanford NLI corpus	Accuracy	78.2	88.3	Chen+ 16

State of the art results on NLP application-level tasks

Task	Test set	Metric	Best non-neural	Best neural	Source
Machine Translation	Enu-deu newstest16	BLEU	31.4	34.8	http://matrix.statmt.org
	Deu-enu newstest16	BLEU	35.9	39.9	http://matrix.statmt.org
Sentiment Analysis	Stanford sentiment bank	5-class Accuracy	71.0	80.7	Socher+ 13
Question Answering	WebQuestions test set	F1	39.9	52.5	Yih+ 15
Entity Linking	Bing Query Entity Linking set	AUC	72.3	78.2	Gao+ 14b
Image Captioning	COCO 2015 challenge	Turing test pass%	25.5	32.2	Fang+ 15
Sentence compression	Google 10K dataset	F1	0.75	0.82	Fillipova+ 15
Response Generation	Sordoni dataset	BLEU-4	3.98	5.82	Li+ 16a

Community

- ACL Anthology (http://www.aclweb.org/anthology/)
- All NLP conference and journal papers, free of charge

ACL Anthology
A Digital Archive of Research Papers in Computational Linguistics

Search the Anthology [ ] via Google | via Searchbench @ DFKI | via AAN @ UMich | via Saffr

The ACL Anthology currently hosts over 33,000 papers on the study of computational linguistics and natural language processing. Subscribe to the mailing list to

NEW The beta version of the new ACL Anthology goes live. It will replace this current version of the Anthology as the default version starting 2015 (don't worry we w

NEW Nov 2014: The December issue of *Computational Linguistics* is now available on the Anthology. Also the Proceedings of the 26th Conference on Computation and the Proceedings of the Australasian Language Technology Association Workshop 2014 (ALTA 2014) are now available on the Anthology. The Proceedings of Language Learning (NLP4CALL) have also been made available on the Anthology.

ACL events

CL:	**Intro FS MT&CL** 74-79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 00 01 02 03 04 05 06 07 08 09 10 11 12 13 UPDATED 14
TACL:	14 13
ACL:	**Intro** 79 80 81 82 83 84* 85 86 87 88 89 90 91 92 93 94 95 96 97* 98* 99 00 01 02 03 04 05 06* 07 08* 09* 10 11 12 13 14
EACL:	**Intro** 83 85 87 89 91 93 95 97* 99 03 06 09 12 14
NAACL:	**Intro** 00* 01 03 04 06* 07* 09* 10* 12* 13*
EMNLP:	96 97 98 99 00 01 02 03 04 05 06 07* 08 09 10 11 12* 13 14
CoNLL:	97 98 99 00 01 02 03 04 05 06 07 08 09 10 11 12 13 14
***Sem/ SemEval:**	98 01 04 07 10 12 13 14

Current NLP Journals

Computational Linguistics

Journal of Natural Language Engineering (JLNE)

Machine Translation

Natural Language and Linguistic Theory

Current NLP conferences

Association for Computational Linguistics

Coling

EACL (Europe Association for Computational Linguistics)

Journal and conferences

- Journal
  - Computational Linguistics
- Conferences
  - ACL: Association for Computational Linguistics (ACL'16 in Berlin!)
  - NAACL: North American Chapter
  - EACL: European Chapter
  - HLT: Human Language Technology
  - EMNLP: Empirical Methods on Natural Language Processing
  - CoLing: Computational Linguistics
  - LREC: Language Resources and Evaluation

End

"The computer is claiming its intelligence is real, and ours is artificial."