
Machine Learning Natural Language

NLP System : IBM Watson

Question Answering System

Quiz show Jeopardy!

- “The first person mentioned by name in 'The man in the Iron mask' is this hero of a previous book by the same author”

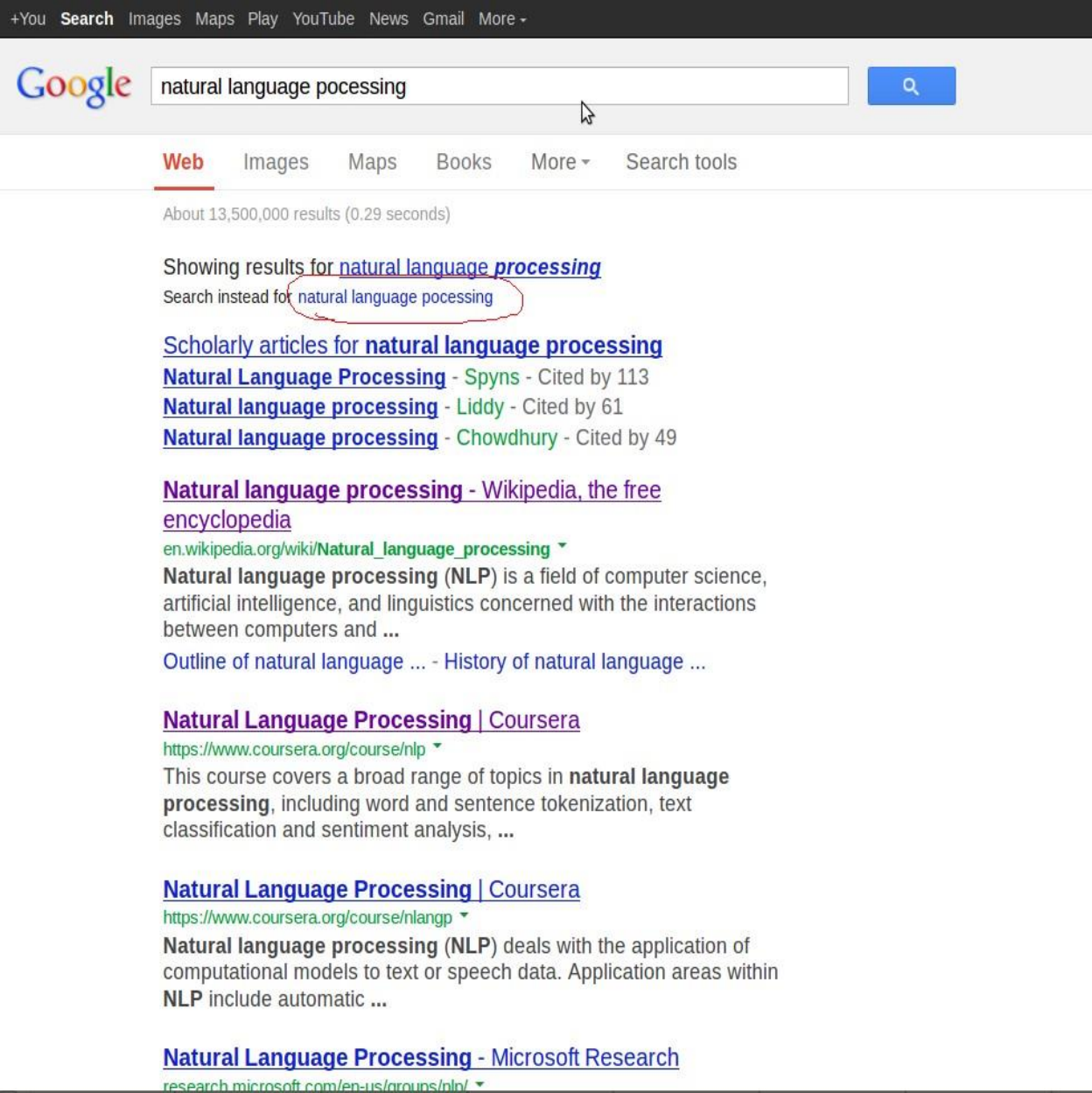


Natural Language Processing

- NLP focuses on developing systems that allow computers to perform useful tasks involving human language
 - Also called Computational Linguistics
- NLP applications
 - Information Retrieval
 - Question Answering
 - Machine Translation
 - Information Extraction

NLP application : Information Retrieval

- Stemming
- Spell checking
- Query expansion
- Word sense disambiguation



The screenshot shows a Google search interface with the query "natural language processing". The search results page displays approximately 13,500,000 results in 0.29 seconds. The results are categorized under the "Web" tab. The first result is a link to "Scholarly articles for natural language processing", which lists three articles: "Natural Language Processing" by Spyns (cited by 113), "Natural language processing" by Liddy (cited by 61), and "Natural language processing" by Chowdhury (cited by 49). The second result is a link to "Natural language processing - Wikipedia, the free encyclopedia", which provides a brief definition of NLP as a field of computer science, artificial intelligence, and linguistics concerned with the interactions between computers and human language. The third result is a link to "Natural Language Processing | Coursera", which describes a course covering a broad range of topics in NLP, including word and sentence tokenization, text classification, and sentiment analysis. The fourth result is another link to "Natural Language Processing | Coursera", which describes a course dealing with the application of computational models to text or speech data. The fifth result is a link to "Natural Language Processing - Microsoft Research".

+You Search Images Maps Play YouTube News Gmail More ▾

Google natural language processing

Web Images Maps Books More ▾ Search tools

About 13,500,000 results (0.29 seconds)

Showing results for natural language processing
Search instead for natural language processing

Scholarly articles for natural language processing

Natural Language Processing - Spyns - Cited by 113
Natural language processing - Liddy - Cited by 61
Natural language processing - Chowdhury - Cited by 49

Natural language processing - Wikipedia, the free encyclopedia
en.wikipedia.org/wiki/Natural_language_processing ▾
Natural language processing (NLP) is a field of computer science, artificial intelligence, and linguistics concerned with the interactions between computers and ...
[Outline of natural language ... - History of natural language ...](#)

Natural Language Processing | Coursera
<https://www.coursera.org/course/nlp> ▾
This course covers a broad range of topics in **natural language processing**, including word and sentence tokenization, text classification and sentiment analysis, ...

Natural Language Processing | Coursera
<https://www.coursera.org/course/nlangp> ▾
Natural language processing (NLP) deals with the application of computational models to text or speech data. Application areas within NLP include automatic ...

Natural Language Processing - Microsoft Research
research.microsoft.com/en-us/groups/nlp/ ▾

NLP application : Question Answering

- Determine type of question and answer
 - Parse the question and identify relations
- POS tagging, Parsing, named entity recognition



The screenshot displays the AnswerBus web application. At the top, the "AnswerBus" logo is centered. Below it, a search bar contains the question "Who is the prime minitser of india ?" (note the typo). To the right of the search bar is an "Ask" button. Below the search bar, a green message reads: "Type in your question! We only want to offer the best possible answers to you." The main content area is titled "Possible Answers" and lists several resources. The first section is "More Resources from Wiki and Web", which includes a link to the "Manmohan Singh (prime minister of India) -- Encyclopedia Britannica" (www.britannica.com/EBchecked/topic/936615/Manmohan-Sing...). Below this, a paragraph describes Manmohan Singh as an Indian economist and politician who became prime minister of India in 2004. Another link is provided: "PMs of India - Prime Minister's Office" (pmindia.nic.in/pmsfindia.php). The next section is "india.gov.in. Prime Ministers of India. Name, Tenure, Party.", which includes a link to "Indira Gandhi (prime minister of India) -- Encyclopedia Britannica" (www.britannica.com/EBchecked/topic/225198/Indira-Gandhi). This is followed by a paragraph about Indira Gandhi as a politician who served as prime minister of India for three consecutive terms (1966-77) and a fourth term from 1980 until her assassination in 1984. A link is provided: "1947-2009: List of Prime Ministers of India - Oneindia News" (news.oneindia.in/2009/05/22/1947-2009-list-of-prime-min...). The final section is "May 22, 2009 ... indian prime ministers, manmohan singh, indian prime minister, ab vajpayee, indira gandhi, lok sabha, jawaharlal nehru, gulzari lal nanda, ...", which includes a link to "Prime Minister of India - NNDB.com" (www.nnadb.com/gov/751/000047610/). At the bottom, a partial link is visible: "Prime Minister of India. GOVERNMENT OFFICE. Prime Minister ...".

NLP application : Machine Translation

- Sentence alignment
- POS tagging
- Parsing
- Sentence generation grammars
- Named Entity Recognition (“New Delhi”)

The screenshot displays the Google Translate interface. At the top, there's a navigation bar with links to various Google services. Below it, the Google logo is visible. A banner promotes downloading Google Chrome. The main section is titled 'Translate' and shows the source language as 'English' and the target language as 'Hindi'. A blue 'Translate' button is present. The input text on the left is: "The history of NLP generally starts in the 1950s, although work can be found from earlier periods. In 1950, Alan Turing published an article titled 'Computing Machinery and Intelligence' which proposed what is now called the Turing test as a criterion of intelligence. This criterion depends on the ability of a computer program to impersonate a human in a real-time written conversation with a human judge, sufficiently well that the judge is unable to distinguish reliably — on the basis of the conversational content alone — between the program and a real human". The output text on the right is the Hindi translation: "काम पहले की अवधि से पाया जा सकता है, हालांकि एनएलपी के इतिहास में आम तौर पर, 1950 के दशक में शुरू होता है. 1950 में एलन ट्यूरिंग अब खुफिया के एक कसौटी के रूप में ट्यूरिंग परीक्षण कहा जाता है का प्रस्ताव है, जो 'कम्प्यूटिंग मशीनरी और खुफिया' शीर्षक से एक लेख प्रकाशित किया. संवादी सामग्री अकेले के आधार पर - - कार्यक्रम और बीच में इस कसौटी भी पर्याप्त न्यायाधीश मजबूती से भेद करने में असमर्थ है कि एक मानव न्यायाधीश के साथ एक वास्तविक समय लिखा बातचीत में एक मानव का रूप धारण करने के लिए एक कंप्यूटर प्रोग्राम की क्षमता पर निर्भर करता है एक वास्तविक मानव". Below the Hindi text, there's a Devanagari script version of the same text.

Try a new browser with automatic translation. [Download Google Chrome](#) [Dismiss](#)

Translate From: English To: Hindi Translate

English Spanish French Detect language

The history of NLP generally starts in the 1950s, although work can be found from earlier periods. In 1950, Alan Turing published an article titled "Computing Machinery and Intelligence" which proposed what is now called the Turing test as a criterion of intelligence. This criterion depends on the ability of a computer program to impersonate a human in a real-time written conversation with a human judge, sufficiently well that the judge is unable to distinguish reliably — on the basis of the conversational content alone — between the program and a real human


English Spanish Hindi


काम पहले की अवधि से पाया जा सकता है, हालांकि एनएलपी के इतिहास में आम तौर पर, 1950 के दशक में शुरू होता है. 1950 में एलन ट्यूरिंग अब खुफिया के एक कसौटी के रूप में ट्यूरिंग परीक्षण कहा जाता है का प्रस्ताव है, जो "कम्प्यूटिंग मशीनरी और खुफिया" शीर्षक से एक लेख प्रकाशित किया. संवादी सामग्री अकेले के आधार पर - - कार्यक्रम और बीच में इस कसौटी भी पर्याप्त न्यायाधीश मजबूती से भेद करने में असमर्थ है कि एक मानव न्यायाधीश के साथ एक वास्तविक समय लिखा बातचीत में एक मानव का रूप धारण करने के लिए एक कंप्यूटर प्रोग्राम की क्षमता पर निर्भर करता है एक वास्तविक मानव


Kāma pahalē kī avadhi sē pāyā jā sakatā hai, hālānki ēna'ēlapī kē itihāsa mēm āma taura para, 1950 kē dasaka mēm śurū hōtā hai. 1950 Mēm ēlana tyūriṅga aba khuphiyā kē ēka kasautī kē rūpa mēm tyūriṅga parīkṣaṇa kahā jātā hai kā prastāva hai, jō" kampyūṭiṅga maśīnarī aura khuphiyā" śīrṣaka sē ēka lēkha prakāśita kiyā. Sanvādī sāmagrī akēlē kē ādhāra para -- kāryakrama aura bīca mēm isa kasautī bhī paryāpta n'yāyādhiśa mazabūtī sē bhēda karanē mēm asamartha hai ki ēka mānava n'yāyādhiśa kē sātha ēka vāstavika samaya likhā bātācīta mēm ēka mānava kā rūpa dhāraṇa karanē kē li'ē ēka kampyūṭara prōgrāma kī kṣamatā para nirbhara karatā hai Ēka vāstavika mānava


NLP application : Information Extraction

- Identifying/Extracting specific kinds of information
- Named entities (NEs): person, location, price, product
 - Mohandas Karamchand Gandhi was born in Porbandar, Gujarath
- Coreference resolution: linking pronouns/abbreviations to entities
 - “Indian Institute of Science” \diamond “IISc.”
- Relations: <DOB>, <spouse>, <attribute>

**Narendra Modi**
Prime Minister of India :
[Overview](#) [History](#) [Education](#) [Movies](#)

**Narendra Modi**
<https://www.narendramodi.in> :
Narendra Modi | NarendraModi.in Official Website of Prime ...
Shri **Narendra Modi**, Get all the information news, updates, speeches on the official website of Prime Minister of India.

**Wikipedia**
https://en.wikipedia.org/wiki/Narendra_Modi :
Narendra Modi
Narendra Damodardas Modi is an Indian politician serving as the 14th and current prime minister of India since May 2014. Modi was the chief minister of ...
Vice President: [Mohammad Hamid Ansari](#); M. ... Alma mater: [University of Delhi \(BA\)](#); Gujar...
President: [Pranab Mukherjee](#); [Ram Nath Kovi...](#) Residence(s): [7, Lok Kalyan Marg, New De...](#)
[PM Narendra Modi](#) · [Jashodaben Modi](#) · [Modi \(disambiguation\)](#) · [Sumitra Mahajan](#)

**Prime Minister of India**
<http://www.pmindia.gov.in> > ... :
Prime Minister of India

About
[narendramodi.in](#)
Narendra Damodardas Modi is an Indian politician serving as the 14th and current prime minister of India since May 2014. Modi was the chief minister of Gujarat from 2001 to 2014 and is the Member of Parliament from Varanasi.
[Wikipedia](#)
Born: 17 September 1950 (age 72 years), [Vadnagar](#)
Spouse: [Jashodaben Modi](#) (m. 1968)
Awards: [Order of the Distinguished Rule of Izzuddin](#), [Seoul Peace Prize](#), [CNN-IBN Indian of the Year](#)
Full name: Narendra Damodardas Modi
Movies: [100 Years of Chrysostom](#)
Siblings: [Amrit Modi](#)
Party: [Bharatiya Janata Party](#)

NLP application : Categorization

- Topical : politics, sports, business
 - Sentiment: positive, negative, neutral
- POS tagging to obtain adjectives

News

India edition Modern

Top Stories

News near you

India

World

Business

Mahindra Satyam

Air India

Ashok Leyland

Indira Gandhi International Airport

Reliance Industries

HCL Enterprise

Emergency landing Mozambique

Pranab Mukherjee India

Technology

Entertainment

Snorts

Business

Nifty holds on to 5100 as Spain surges 5%

NDTV - 10 minutes ago

A strong opening on European bourses, particularly Spain, helped Indian stocks maintain early gains Monday. The BSE Sensex traded nearly 150 points higher, while the Nifty index managed to stay above the 5100 mark post noon.

Sensex touches month high; Nifty holds 5100 Economic Times

Markets hold onto gains in late noon deals Business Standard

See all 36 sources »

5 things the numbers tell you

NDTV - 5 minutes ago

Car sales in India rose an annual 2.8 per cent in May, according to data from the Society of Indian Automobile Manufacturers, a seventh consecutive monthly increase but far below industry expectations, with demand hit by a hike in excise duty on the ...

Govt empowers Air India to sack 300 pilots on strike

Hindustan Times - 9 minutes ago

With Air India considering further crackdown on the striking pilots, the government today said it



Canon PowerShot SX40 HS 12.1 MP Digital Camera

\$330 online

★★★★★ 524 reviews +19 Recommend this on Google

#1 in Digital Cameras

September 2011 - Canon - Point & Shoot - 12.1 megapixel - Electronic Viewfinder - Compact - CMOS - Pop-up Flash - ISO 3200

« Back to overview

Reviews

Summary - Based on 524 reviews



What people are saying

pictures	★★★★★	"Picture clarity is great."
features	★★★★★	"Product has a lot of features for the price."
zoom/lens	★★★★★	"Good if not ready for DSLR."
design	★★★★★	"Another point is the overall camera speed."
video	★★★★★	"Great video quality."
screen	★★★★★	"Flip out screen is handy."
battery life	★★★★★	"The battery life is pretty good."

Canon PowerShot SX40 HS Review

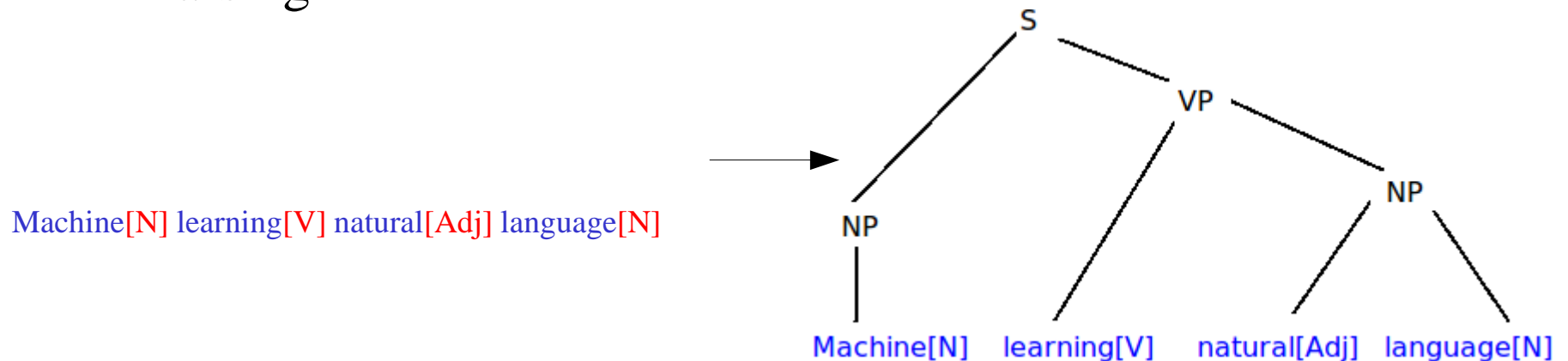
★★★★★ By ConsumerSearch - Oct 31, 2011 - Editorial review - ConsumerSearch

Pros: Versatile lens range, excellent CMOS image sensor, shoots full 1080p HD video, optical image stabilization, lots of manual control, improved shooting speed

Cons: No RAW mode, relatively small (2.7-inch) LCD screen

NLP : Tasks

- Segmentation : words, sentences
- Morphology : plural “boy” “boys”, “agree” ---> “agreement”
Stemming “fishing”, “fished”, “fish”, “fisher” ---> “fish”
- Syntactic Analysis : structural relationships between words
 - Part of Speech (POS) Tagging
Machine[N] learning[V] natural[Adj] language[N]
 - Parsing



NLP : Tasks

- Semantics

- Word Sense Disambiguation : “I went to *bank*”
- Semantic role labelling :
“*Mary*[Agent] sold the book[goods] to *John*[Receipient]”

- Pragmatics : how language is used to accomplish goals

- I’m sorry Dave, I’m afraid I can’t do that [Polite]
- I can't do that [Rude]

- Discourse

Coreference Resolution : linking pronouns/abbreviations to entities

“I saw *Scott* yesterday. *He* was fishing by the lake.”

“*Indian Institute of Technology Hyderabad* is a public institution located in Hyderabad. *IITH* was established in 2007.”

Named Entity recognition (NER) : person, location, price, product

Mohandas Karamchand Gandhi was born in *Porbandar*, *Gujarath*

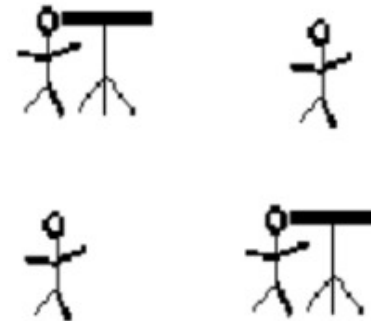
NLP is hard

- Natural language is **ambiguous**
- Sentence Segmentation : “I went out with Mr. Smith.”
- Syntactic

“Flies[Noun/Verb] like flower[Noun/Verb]”

“I saw the man with the telescope” vs

“I saw the man with the telescope”

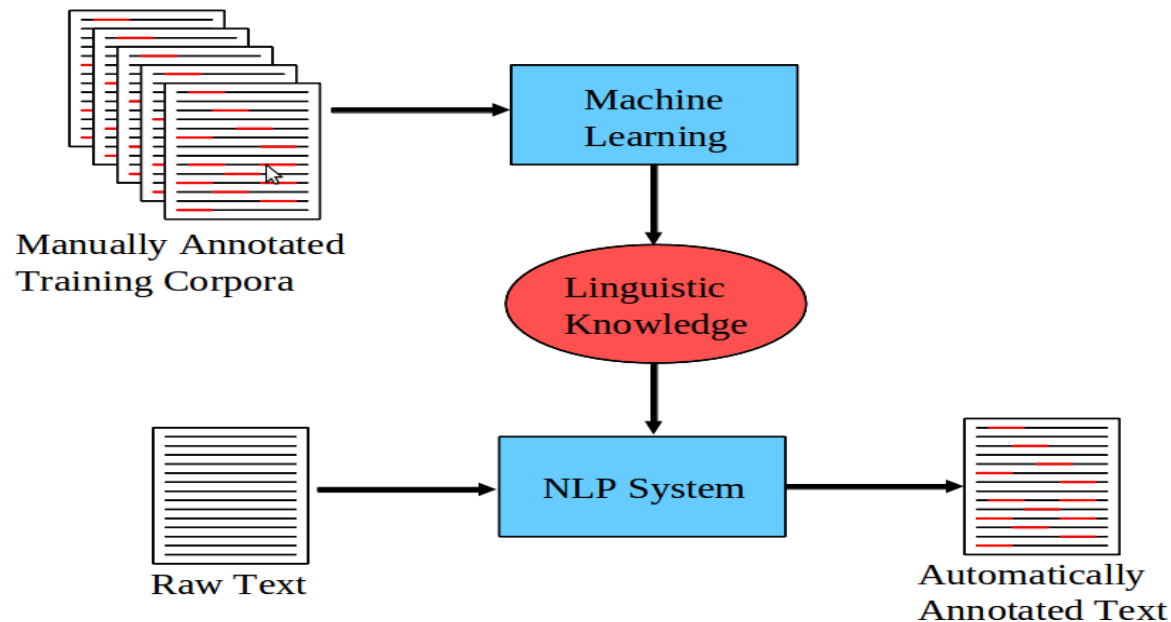


- Semantic
“I put the **plant** in the window” vs “Ford put the **plant** in Mexico”
- Ambiguity is Explosive
“I saw the man on the hill with the telescope.”: 4 parses



Machine Learning Natural Language

- “Rules” in language have numerous exceptions and irregularities
- Manual knowledge engineering, is difficult, time-consuming, and error prone.
- Use machine learning methods to automatically acquire the required knowledge from appropriately annotated text corpora.
- Annotating corpora is easier and requires less expertise than manual knowledge engineering.



Machine Learning POS Tagging

- Lowest level of syntactic analysis
- Useful for Parsing and word sense disambiguation
- Ambiguity in POS tagging

Flies[Noun] like[Verb] flower[Noun]

Time flies[Verb] like[Prep] an arrow.

Learning : Train models on human annotated corpora like the Penn Treebank.

1 Pierre/NNP Vinken/NNP ,/, 61/CD years/NNS old/JJ ,/, will/MD join/VB the/DT board/NN as/IN a/DT nonexecutive/JJ director/NN Nov./NNP 29/CD ./.

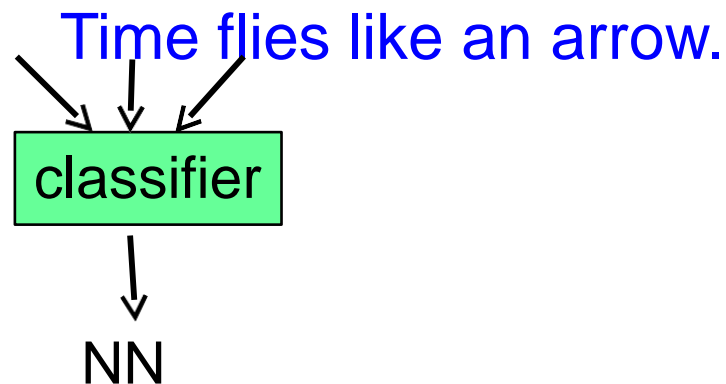
2 Mr./NNP Vinken/NNP is/VBZ chairman/NN of/IN Elsevier/NNP N.V./NNP ,/, the/DT Dutch/NNP publishing/VBG group/NN ./.

3 Rudolph/NNP Agnew/NNP ,/, 55/CD years/NNS old/JJ and/CC chairman/NN of/IN Consolidated/NNP Gold/NNP Fields/NNP PLC/NNP ,/, was/VBD named/VBN a/DT nonexecutive/JJ director/NN of/IN this/DT British/JJ industrial/JJ conglomerate/NN ./.

POS Tagging

Classification

Classify each word independently but use as input features, information about the surrounding words.

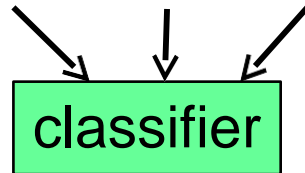


POS Tagging

Classification

NN

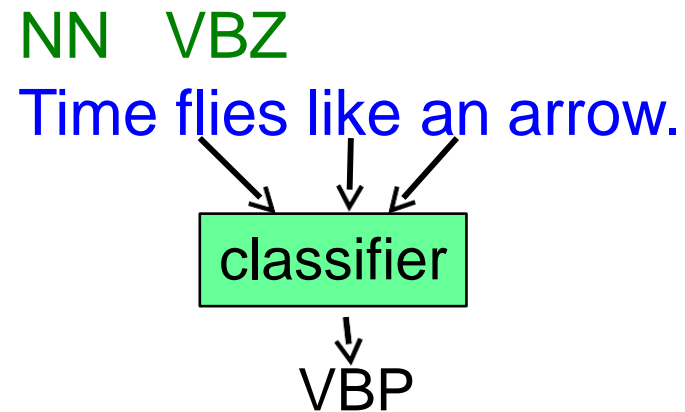
Time flies like an arrow.



↓
VBZ

POS Tagging

.Classification

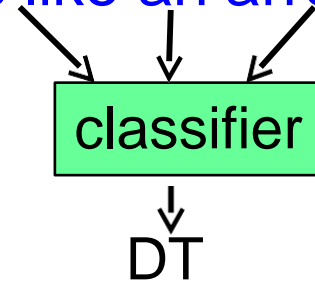


POS Tagging

.Classification

NN VBZ VBP

Time flies like an arrow.

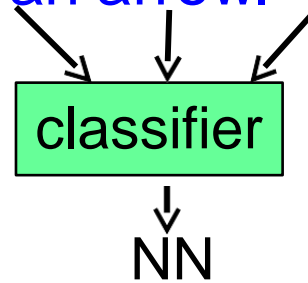


POS Tagging

Classification

NN VBZ VBP DT

Time flies like an arrow.



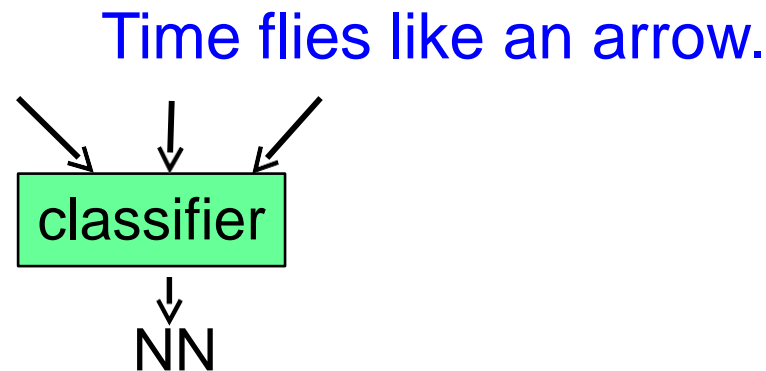
POS Tagging

Classification

NN VBZ VBP DT NN
Time flies like an arrow.

Sequence Labeling

Tags of words are dependent on the tags of other words in the sentence, particularly their neighbors

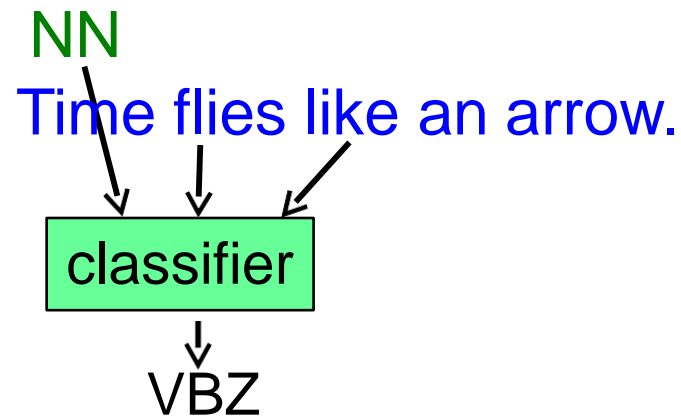


POS Tagging

Classification

NN VBZ VBP DT NN
Time flies like an arrow.

Sequence Labeling

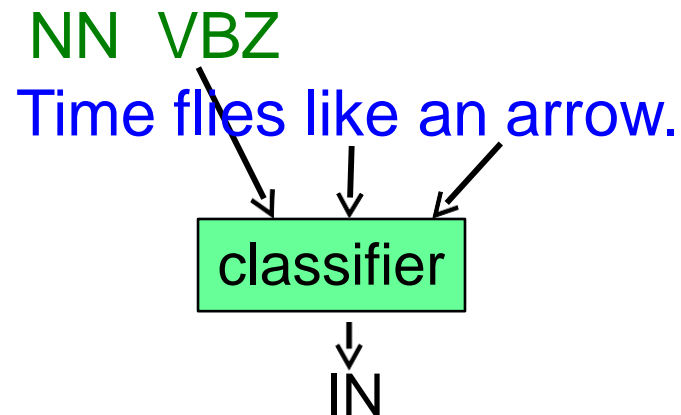


POS Tagging

Classification

NN VBZ VBP DT NN
Time flies like an arrow.

Sequence Labeling

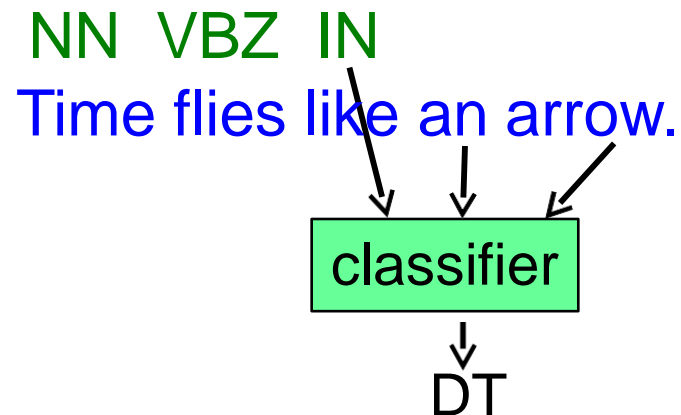


POS Tagging

Classification

NN VBZ VBP DT NN
Time flies like an arrow.

Sequence Labeling



POS Tagging

Classification

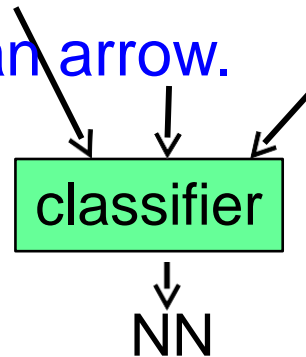
NN VBZ VBP DT NN

Time flies like an arrow.

Sequence Labeling

NN VBZ IN DT

Time flies like an arrow.



Sequence Labeling

Classification

NN VBZ VBP DT NN
Time flies like an arrow.

Sequence Labeling

NN VBZ IN DT NN
Time flies like an arrow.

POS Tagging is best modeled as a **sequence learning problem** than as a classification problem

- Information Extraction, Named Entity recognition

Statistical models: Hidden Markov Model (HMM), Maximum Entropy Markov Model (MEMM), Conditional Random Field (CRF)

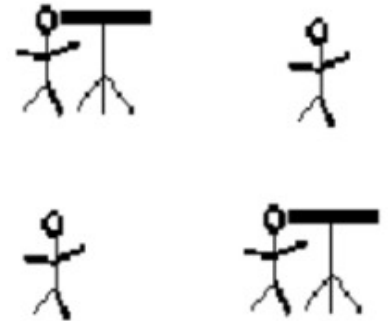
Parsing

- Ambiguity

“I saw the man with the telescope” vs

“I saw the man with the telescope”

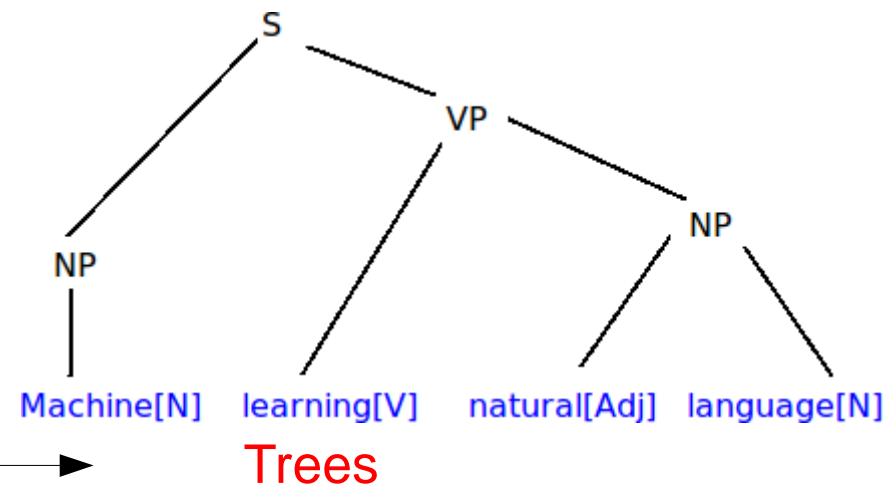
Probabilistic Context Free Grammars (PCFG)



- Structured Prediction

Machine learning natural language

Strings



Statistical models: Conditional Random Field, Structured perceptrons, Structured support vector machines

Machine learning for NLP

- Transfer Learning, domain adaptation
 - Adapting a model learned on a resource rich language to resource scarce language
- Deep learning
 - Unsupervised learning of useful features
- Conferences : Association of Computational Linguistics(ACL), Computational Linguistics (COLING), Empirical Methods in NLP (EMNLP)
- Software tools
 - Stanford CoreNLP, openNLP, NLTK, Lingpipe

References

Daniel Jurafsky and James H. Martin (2008). Speech and Language Processing

Christopher D. Manning and Hinrich Schütze (1999). Foundations of Statistical Natural Language Processing.

Machine Learning Methods in Natural Language Processing

http://www.cs.columbia.edu/~mcollins/papers/tutorial_colt.pdf

Lafferty, J., McCallum, A., Pereira, F. (2001). Conditional random fields: Probabilistic models for segmenting and labeling sequence data.

Ioannis Tsochantaridis, Thorsten Joachims, Thomas Hofmann and Yasemin Altun (2005), Large Margin Methods for Structured and Interdependent Output Variables

Deep learning for NLP,

<http://www.socher.org/index.php/DeepLearningTutorial/DeepLearningTutorial>

Thank you