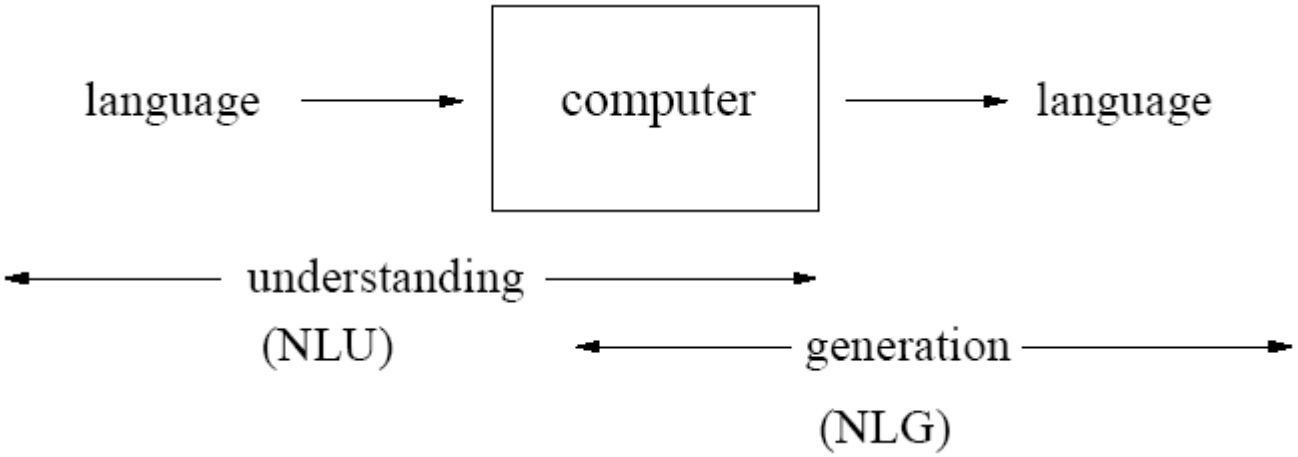
Introduction to Natural Language Processing

(Some slides adapted from Ralph Grishman at NYU,
Yejin Choi at UWashington, N. Tomura at Depaul, Jurafsky and Martin,
CS224N, CS224d at Stanford and other resourses on the web)

What is NLP?

- Natural Language Processing (NLP) is a field in Artificial Intelligence (AI) devoted to creating computers that use natural language as input and/or output.



- An interdisciplinary field with many names corresponding to its many facets, names like speech and language processing, human language technology, natural language processing, computational linguistics, and speech recognition and synthesis.

What is NLP?

- Natural language processing is a field at the intersection of
  - computer science
  - artificial intelligence
  - and linguistics.
- Goal: for computers to process or "understand" natural language in order to perform tasks that are useful, e.g.
  - Question Answering
- Fully **understanding and representing** the **meaning** of language (or even defining it) is an illusive goal.
- Perfect language understanding is AI-complete

(*) To call a problem AI-complete reflects an attitude that it would not be solved by a simple specific algorithm.

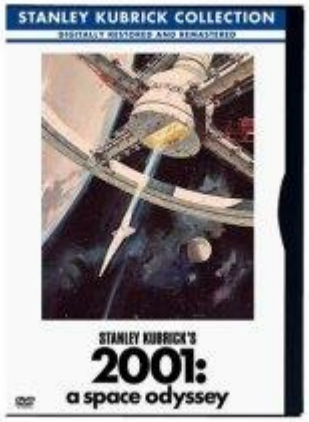
Source: cs224d

Why NLP?

- The goal of this new field is to get computers to perform useful tasks involving human language, tasks like enabling human-machine communication, improving human-human communication, or simply doing useful processing of text or speech.
- To interact with computing devices using human (natural) languages. For example,
  - Building intelligent robots (AI).
  - Enabling voice-controlled operation.
- To access (large amount of) information and knowledge stored in the form of human languages quickly.

Early days of NLP: The Dream: Machines that Can Speak

- One example of a useful such task is a **conversational agent**.
- HAL 9000 in "2001: A Space Odyssey" (*)



(*) The HAL 9000 computer in Stanley Kubrick's film *2001: A Space Odyssey* is one of the most recognizable characters in 20th century cinema. HAL is an artificial agent capable of such advanced language behavior as speaking and understanding English, and at a crucial moment in the plot, even reading lips.

Machines that Can Speak (cont.)

- C3PO
  in Star Wars

- KITT
  in Knight Rider

But Still a Sci-Fi…

- "However, as 2001 approached it became clear that *2001*'s predictions in computer technology were far fetched. Natural language, lip reading, planning and plain common sense in computers were still the stuff of science fiction."
  HAL 9000 - Wikipedia

Test for Intelligence – Turing Test

- Alan Turing (1950) proposed a test of a machine's capability to perform human-like conversation.

  - A human judge engages in a natural language conversation with two other parties, one a human and the other a machine; if the judge cannot reliably tell which is which, then the machine is said to pass the test.

Early Conversational Programs

- ELIZA (by Joseph Weizenbaum), 1966
  - A psychotherapist
  - No real understanding; simple pattern-matching to respond to user input -> canned responses

(my ?x depresses me) (why does your ?x depress you) 5

(life ?x) (why do you say it ?x) 3

(I could ?x) (you could ?x) 2

(because ?x) (that is a good reason) 3

(?x) (tell me more) 0

>> Hello.
How do you do. Please state your problem.
>> Men are all alike.
In what way ?
>> They're alwas bugging us about something or other.
I'm not sure I understand you fully.
>> They're always bugging us about something or other.
Can you think of a specific example ?
>> Well, my boyfriend made me come here.
Your boyfriend made you come here ?
>> He says I'm depressed much of the time.
I am sorry to hear that you are depressed.

Dialogue systems/conversational agents

- ELIZA and HAL examples of dialogue systems
- Require both understanding and generation
  - Dave: Open the pod bay doors, HAL.
  - HAL: I'm sorry Dave, I'm afraid I can't do that.
  - Dave: What's the problem?
  - HAL: I think you know what the problem is just as well as I do.

Modern NLP:
NLP in the Days of Big Data

Three trends:

1. An **enormous amount of information** is now available in machine readable form as natural language text (newspapers, web pages, medical records, financial filings, product reviews, discussion forums, etc.)
2. Conversational agents are becoming an important form of human-computer **communication**
3. Much of human-human interaction is now mediated by computers via **social media**

NLP Applications

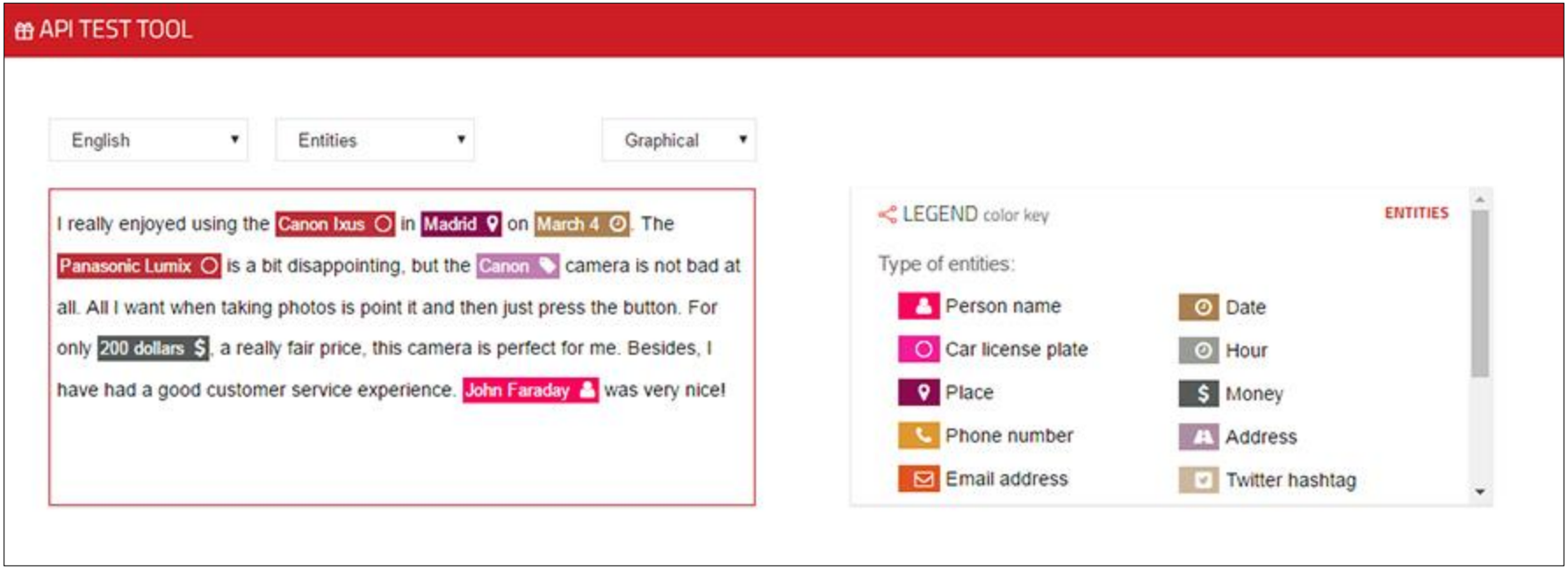
- Three prominent application areas:
  - Text analytics/mining (from *"**unstructured data**")*
    - Sentiment analysis
    - Topic identification
    - Digital Humanities (*"new ways of doing scholarship that involve collaborative, transdisciplinary, and computationally engaged research, teaching, and publishing."*)
  - Conversational agents
    - Siri, Cortana, Amazon Alexa, Google Assistant
    - Chatbots
  - Machine translation

Text Analytics

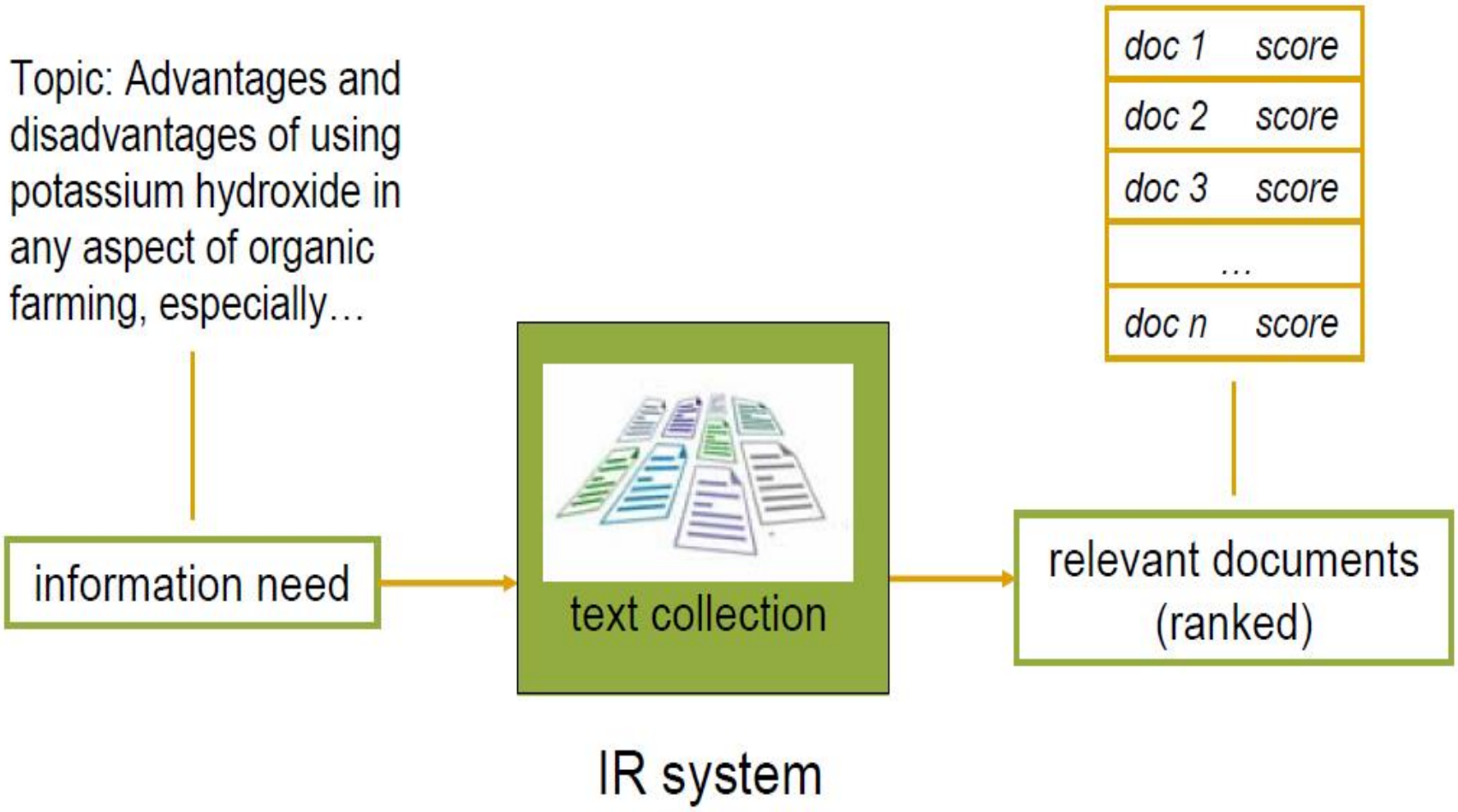
- Data-mining of weblogs, microblogs, discussion forums, user reviews, and other forms of user-generated media.

Text Analytics (cont.)

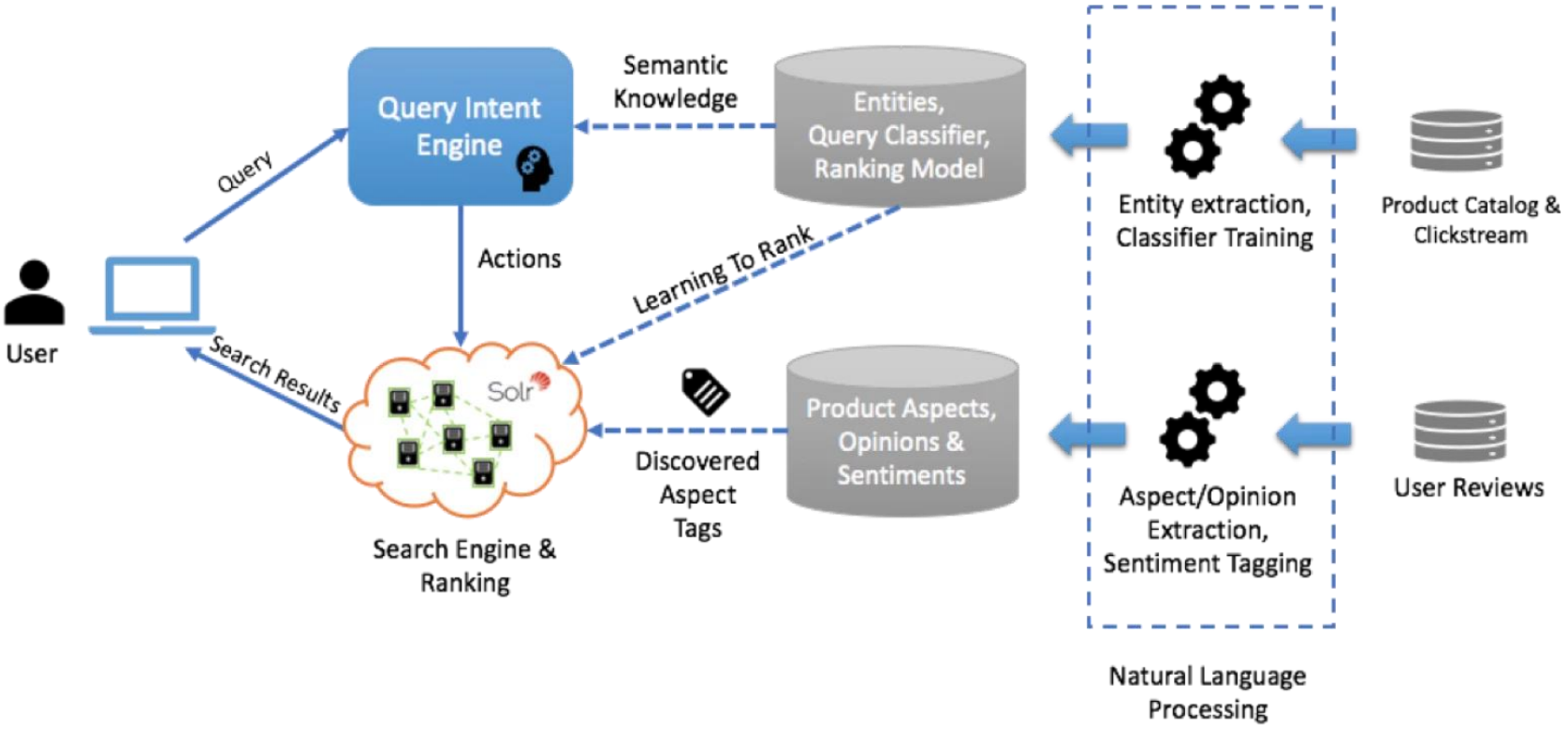
- Typically this involves the extraction of **limited** kinds of semantic and pragmatic information from texts
  - Entity mentions
  - Concept identification
  - Sentiment



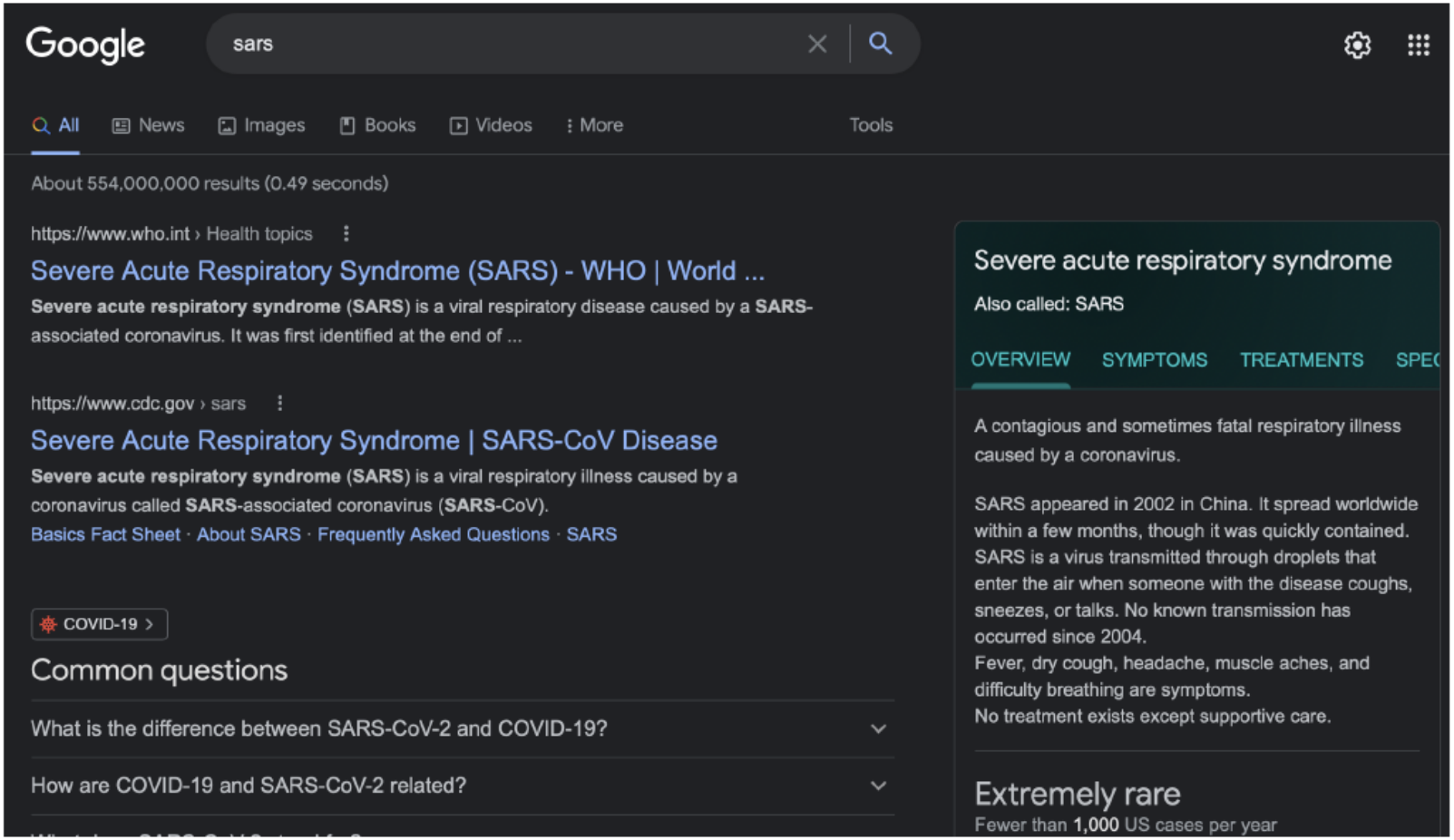
Information retrieval



Use Machine Learning and NLP to Empower Search



Search, and way beyond search



Information Extraction

- **Information extraction is the process of extracting information from unstructured textual sources to enable finding entities as well as classifying and storing them in a database.**
- **Ex: Convert 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

Text Summarization using NLP

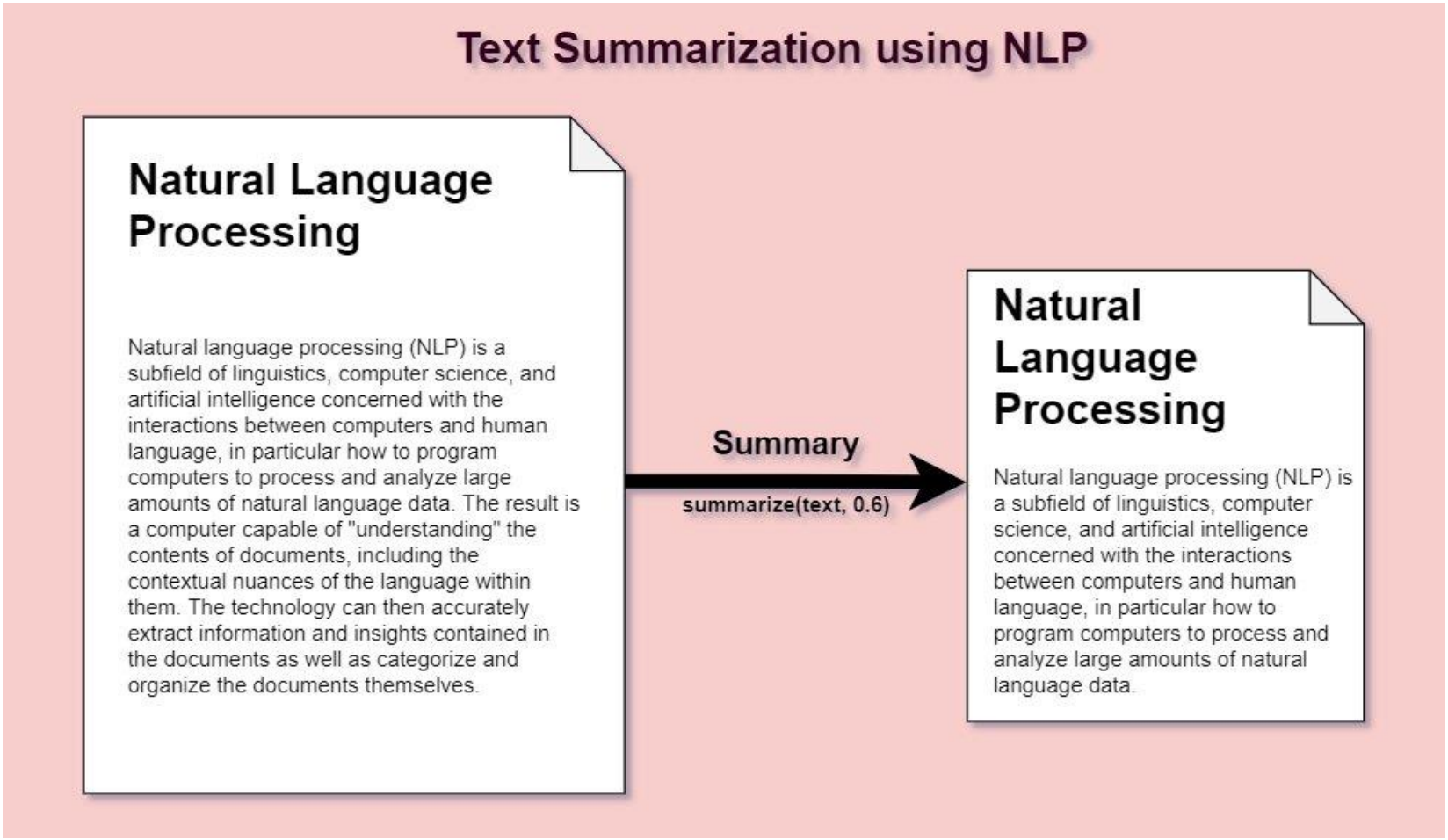


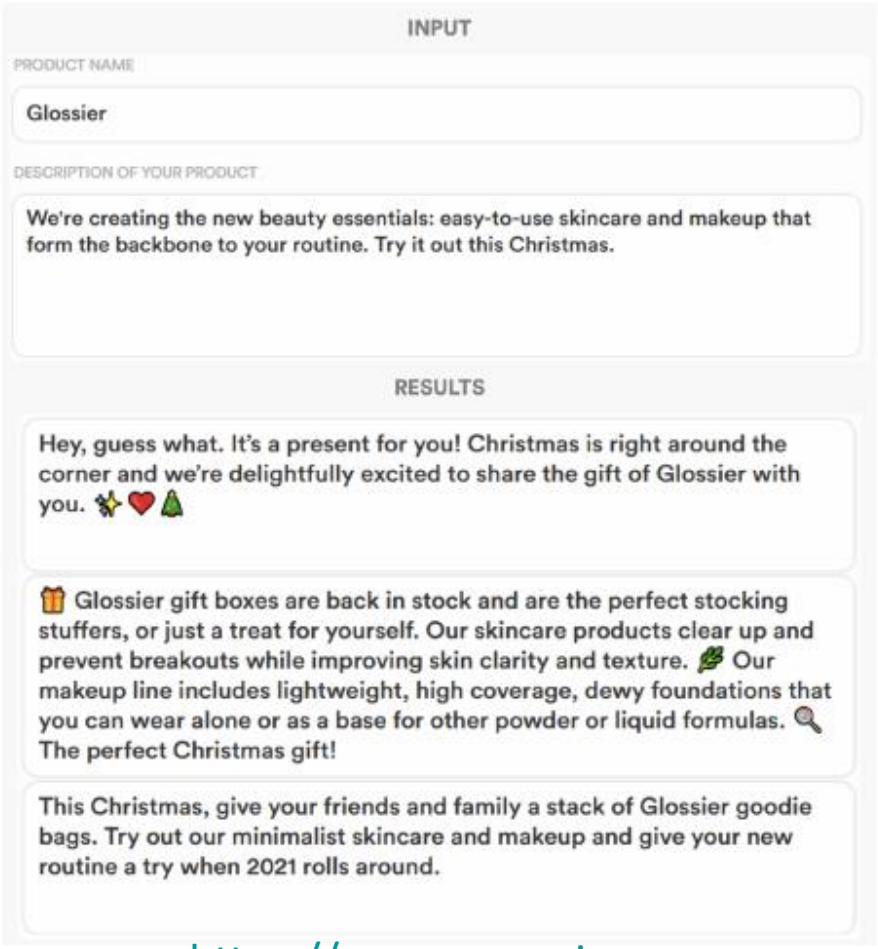
Image captioning

Sutskever et al. 2014

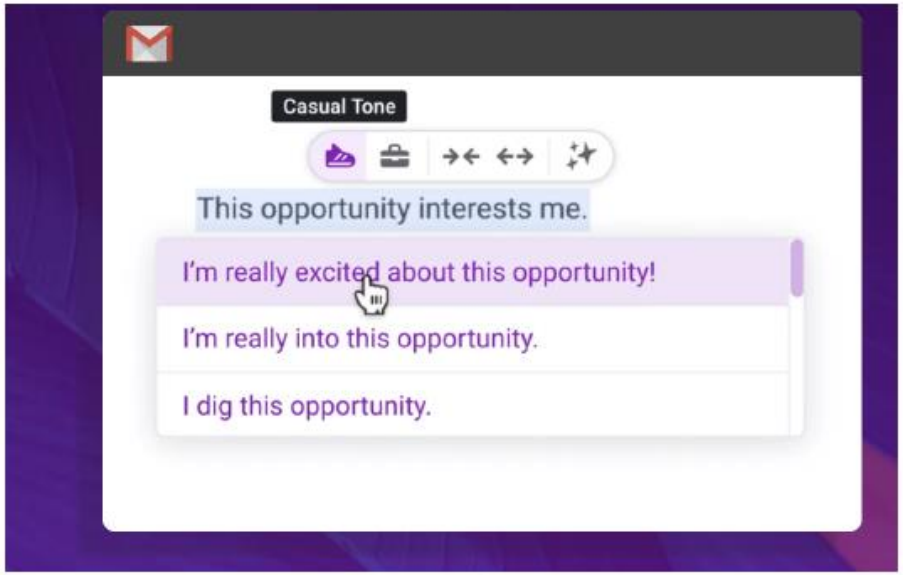
Demo

- Sentiment Analysis with Python NLTK Text Classification
  - http://text-processing.com/demo/sentiment/
- Tweet Sentiment Visualization Tool
  - https://www.csc2.ncsu.edu/faculty/healey/tweet_viz/tweet_app/
- Concept Extraction
  - http://aylien.com/
  - https://www.textrazor.com/
- Text Summarization
  - https://quillbot.com/summarize
  - https://www.paraphraser.io/text-summarizer
  - https://www.summarizer.org/

Demo: Text generation



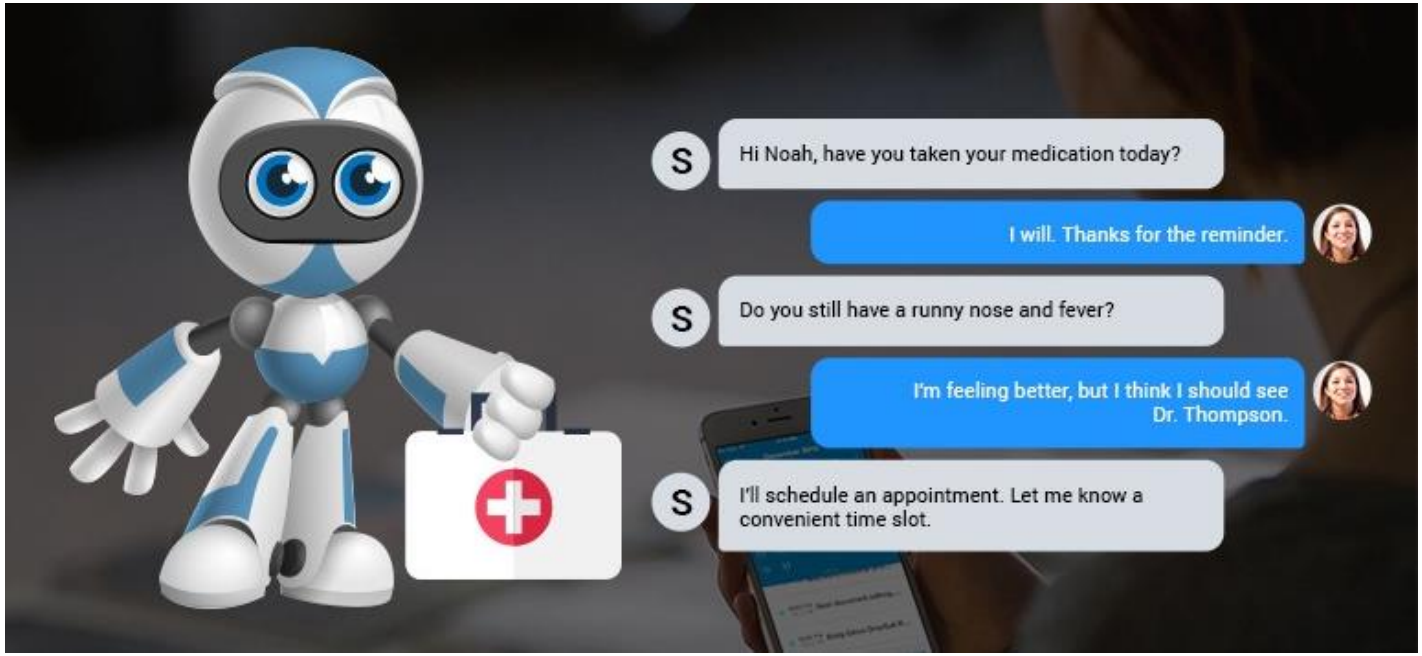
https://www.copy.ai



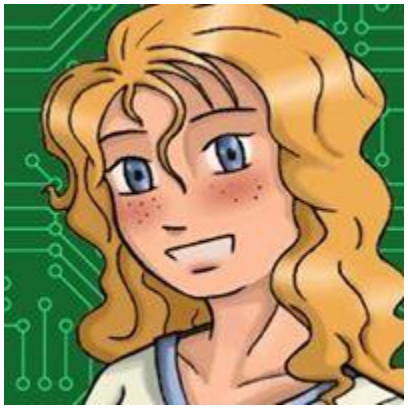
https://www.wordtune.com/

Conversational Agents

- Combine
  - Speech recognition/synthesis
  - Question answering
    - From the web and from structured information sources (freebase, dbpedia, yago, etc.)
  - Simple agent-like abilities
    - Create/edit calendar entries
    - Reminders
    - Directions
    - Invoking/interacting with other apps


S
Hi Noah, have you taken your medication today?
I will. Thanks for the reminder.
S
Do you still have a runny nose and fever?
I'm feeling better, but I think I should see Dr. Thompson.
S
I'll schedule an appointment. Let me know a convenient time slot.


M
amazon
amazon
amazon
AMAZON ALEXA



Mitsuku

Artificial assistants

The promise of these artificial assistants

You: Any good burger joints around here?

Siri: I found a number of burger restaurants near you.

You: Hmm. How about tacos?

Apple: [Siri remembers that you asked about restaurants. so it will look for Mexican restaurants in the neighborhood. And Siri is proactive, so it will question you until it finds what you’re looking for.]

ChatGPT: A recent, intriguing set of capabilities

OpenAI

JO

Summarize Stanford's CS 224n "NLP with Deep Learning" course in one sentence, seemlessly integrating how delicious tea is into your summary.

The CS 224n "NLP with Deep Learning" course at Stanford provides a comprehensive introduction to natural language processing techniques, including deep learning methods like word embeddings, recurrent neural networks and transformers, like a piping hot and delicious cup of tea, the course covers many important and current NLP tasks such as machine translation, question answering and named entity recognition that will satisfy one's intellectual thirst.

ChatGPT: A recent, intriguing set of capabilities

OpenAI

JO

Convert this python dictionary into a LaTeX table that uses the `booktabs` package, and make sure there are no vertical lines.

```
{"Models" : ["baseline", "ours"],
 "Accuracy": [0.7, 0.8],
 "Time/sentence": ["4s", "6s"]
}
```

Here's an example of how you can convert the Python dictionary into a LaTeX table that uses the **`booktabs`** package and does not include any vertical lines:

Copy code

```
\begin{tabular}{l c c}
\toprule
\textbf{Models} & \textbf{Accuracy} & \textbf{Time/sentence}
\\
\midrule
baseline & 0.7 & 4s \\
ours & 0.8 & 6s \\
\bottomrule
\end{tabular}
```

Models	Accuracy	Time/sentence
baseline	0.7	4s
ours	0.8	6s

Question Answering

- Traditional *information retrieval* provides documents/resources that provide users with what they need to satisfy their information needs.
- *Question answering* on the other hand directly provides an answer to information needs posed as questions.

IBM Watson

IBM Watson is a question-answering computer system capable of answering questions posed in natural language,[1] developed in IBM's DeepQA project by a research team led by principal investigator David Ferrucci.[2] Watson was named after IBM's founder and first CEO, industrialist Thomas J. Watson.[3][4] The computer system was initially developed to answer questions on the quiz show *Jeopardy!*[5] and, in 2011, the Watson computer system competed on *Jeopardy!* against champions Brad Rutter and Ken Jennings,[3][6] winning the first place prize of $1 million.[7]

https://www.youtube.com/watch?v=WFR3lOm_xhE

Question Answering Architecture

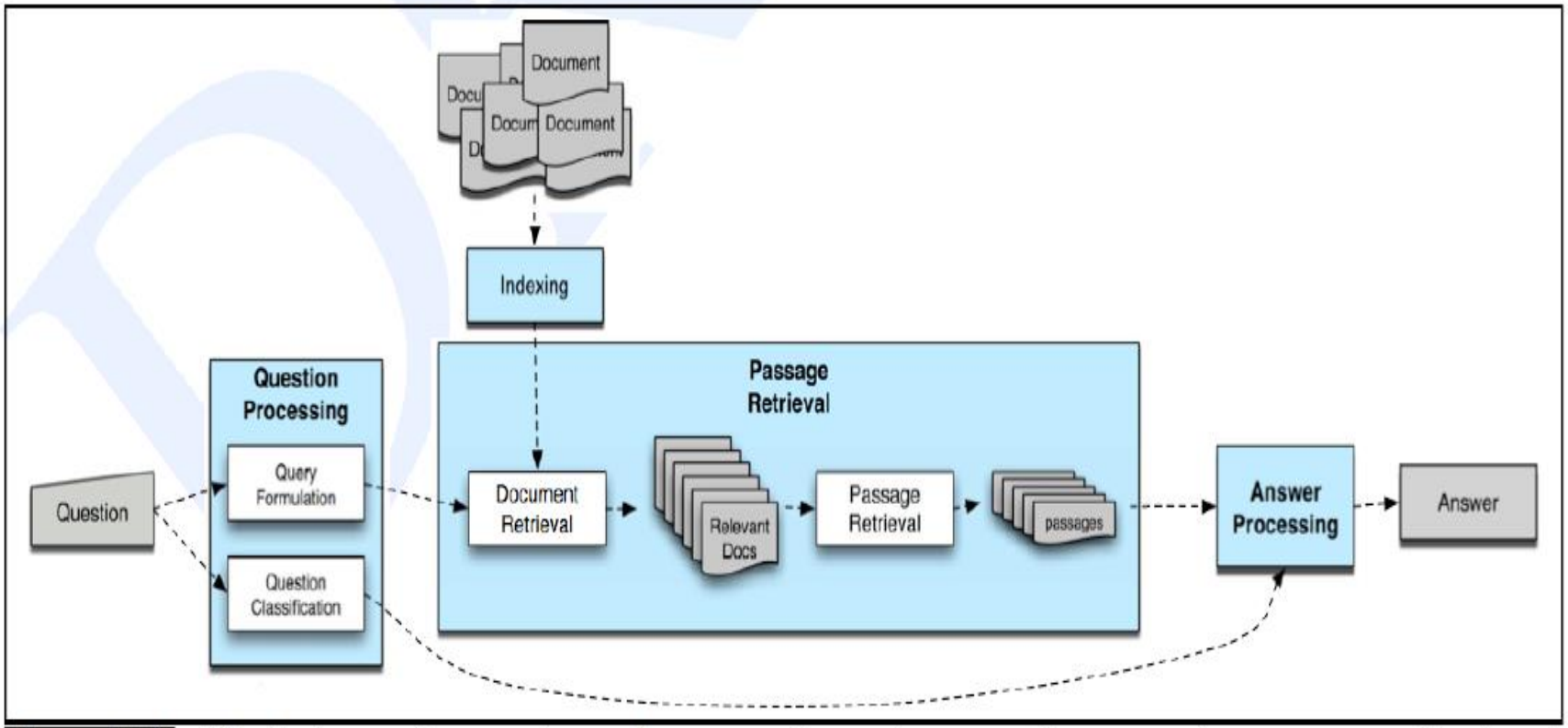
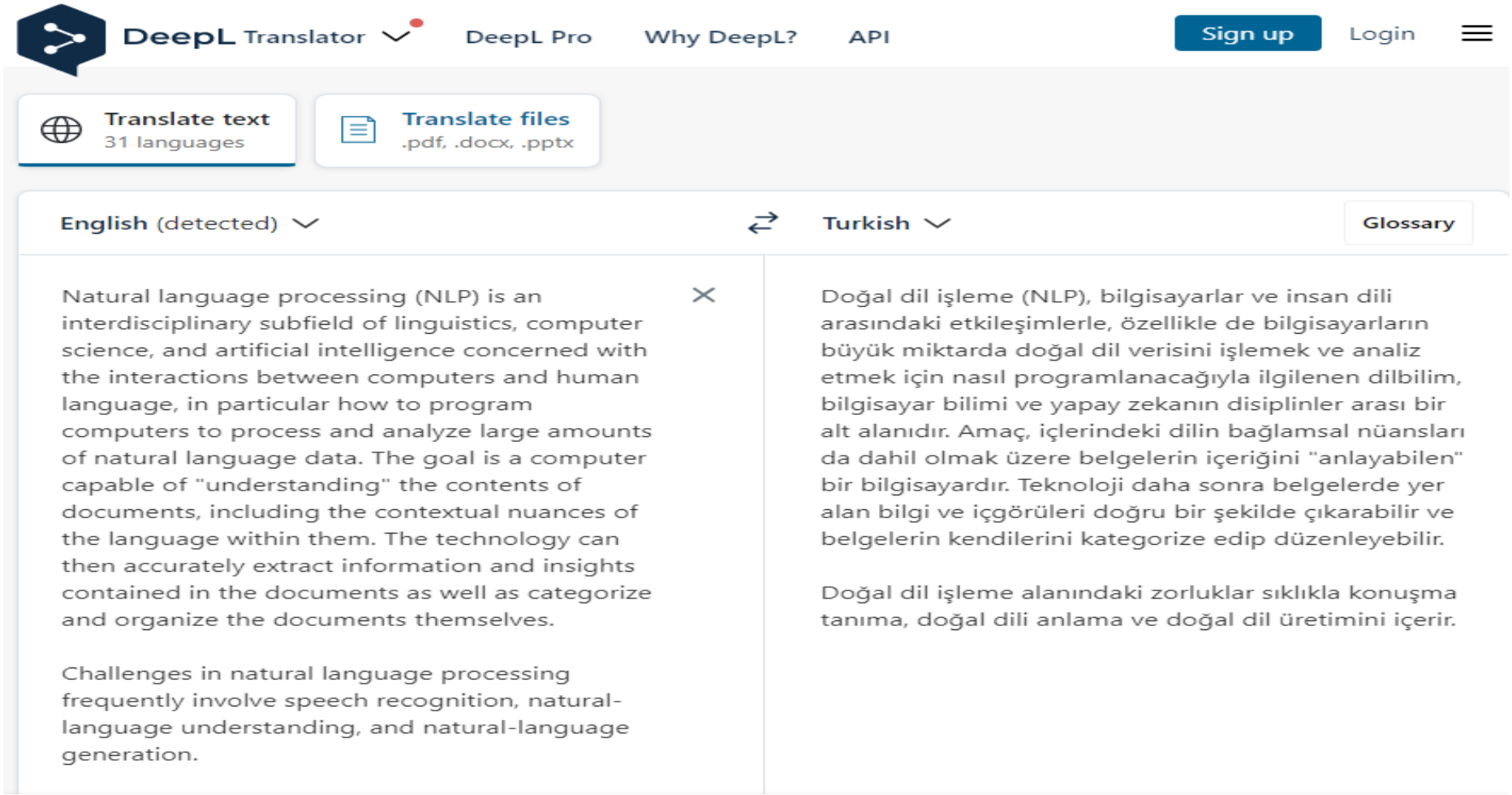


Figure 23.8 The 3 stages of a generic question answering system: question processing, passage retrieval, and answer processing..

Machine Translation

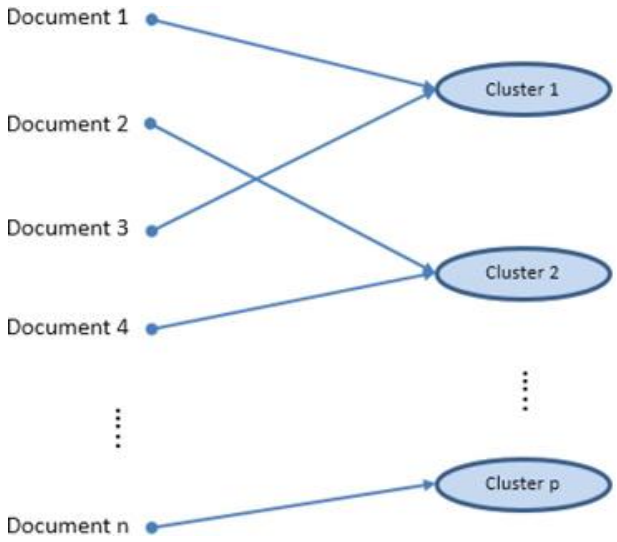
- The automatic translation of texts between languages is one of the oldest non-numerical applications in Computer Science.
- In the past 15 years or so, MT has gone from a niche academic curiosity to a robust commercial industry.

Trained on text data, neural machine translation is quite good!



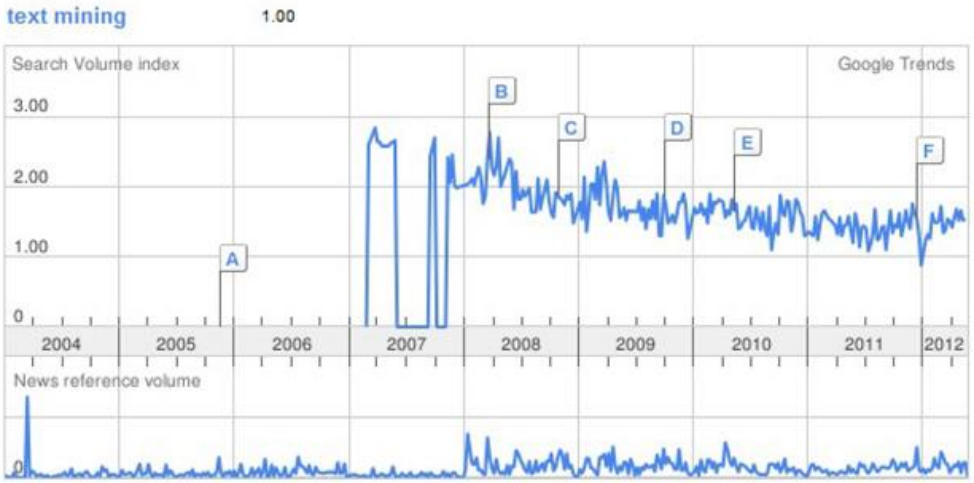
Text Mining Applications – Unsupervised

- Text clustering



Cluster No.	Comment	Key Words
1	1, 3, 4	doctor, staff, friendly, helpful
2	5, 6, 8	treatment, results, time, schedule
3	2, 7	service, clinic, fast

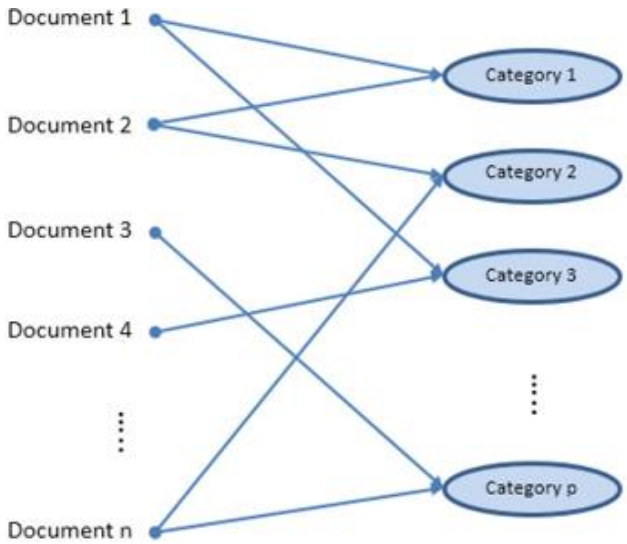
- Trend analysis



Trend for the Term “text mining” from Google Trends

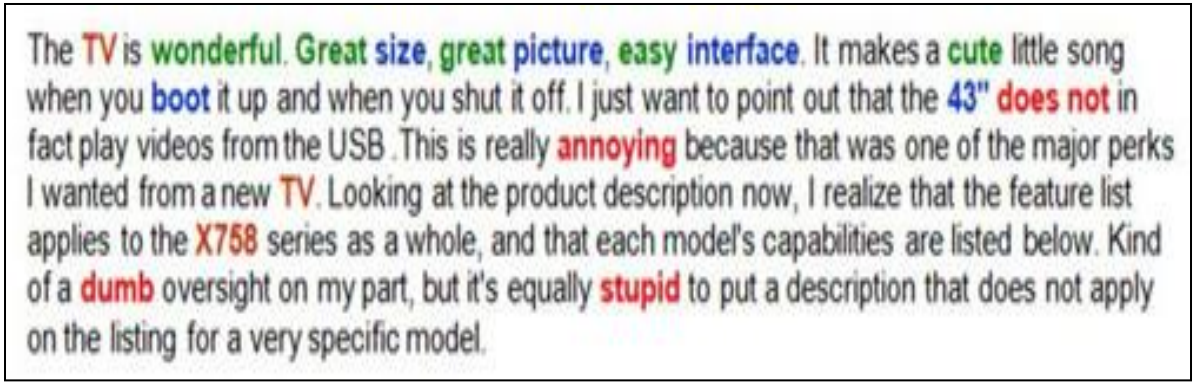
Text Mining Applications – Supervised

- Many typical **predictive modeling** or classification applications can be enhanced by incorporating textual data in addition to traditional input variables.
  - churning propensity models that include customer center notes, website forms, e-mails, and Twitter messages
  - hospital admission prediction models incorporating medical records notes as a new source of information
  - insurance fraud modeling using adjustor notes
  - sentiment categorization (next page)
  - stylometry or forensic applications that identify the author of a particular writing sample



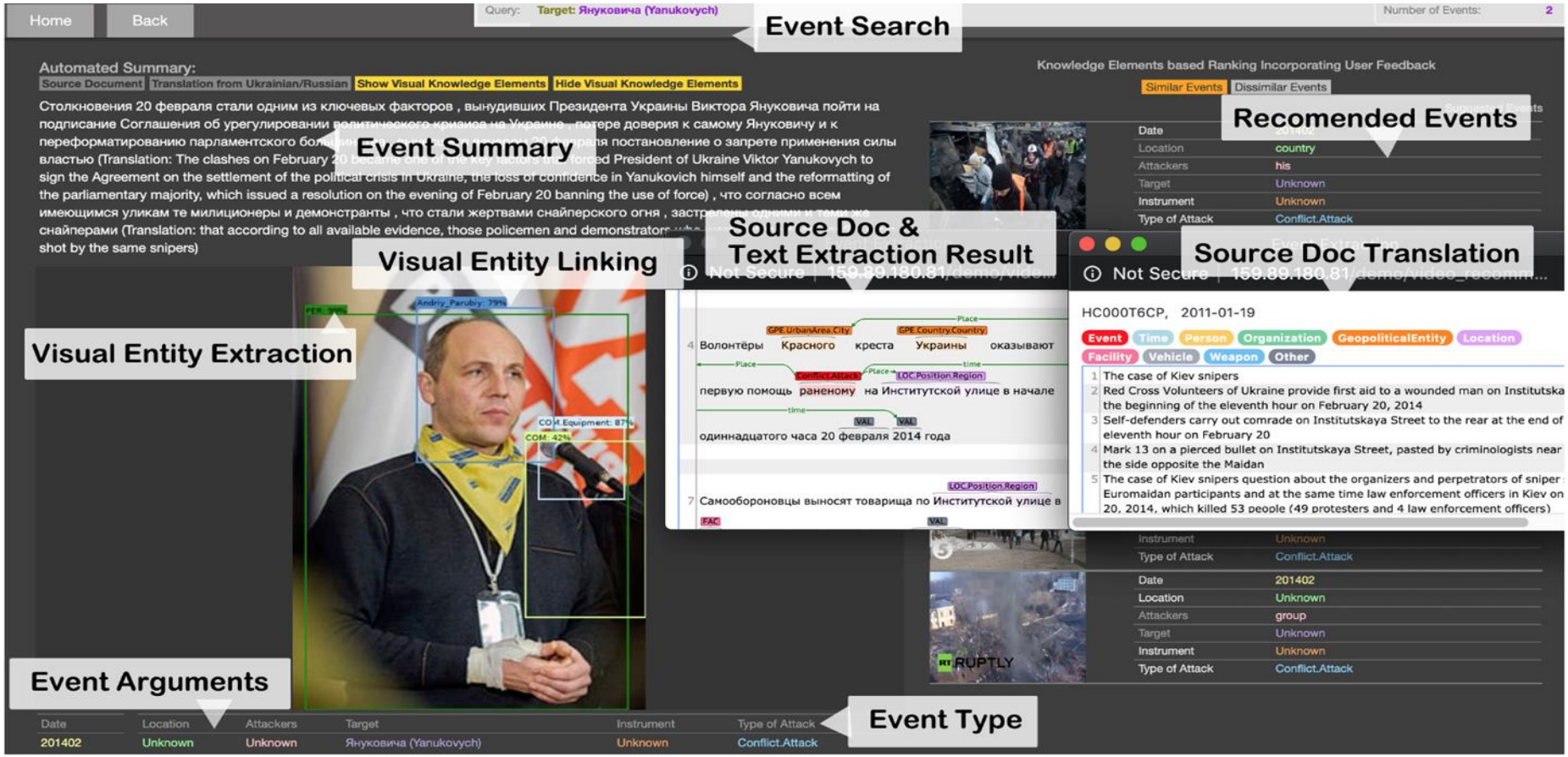
Sentiment Analysis

- The field of sentiment analysis deals with categorization (or classification) of opinions expressed in textual documents



Green color represents positive tone, red color represents negative tone, and product features and model names are highlighted in blue and brown, respectively.

News Understanding and Recommendation (Li et al., ACL2020Demo)



(Li et al., ACL2020 Best Demo Paper Award)
GitHub: https://github.com/GAIA-IE/gaia
DockerHub: https://hub.docker.com/orgs/blendernlp/repositories
Demo: http://159.89.180.81/demo/video_recommendation/index_attack_dark.html

Event-centric Question Answering

- Identifying event orders and predicting future events

Heavy snow is causing disruption to transport across the UK, with heavy rainfall bringing flooding to the south-west of England. Rescuers searching for a woman trapped in a landslide at her home in Looe, Cornwall, said they had found a body.

Q1: What events have already finished?
A: searching trapped landslide said found
Q2: What events have begun but has not finished?
A: snow causing disruption rainfall bringing flooding
Q3: What will happen in the future?
A: No answers.

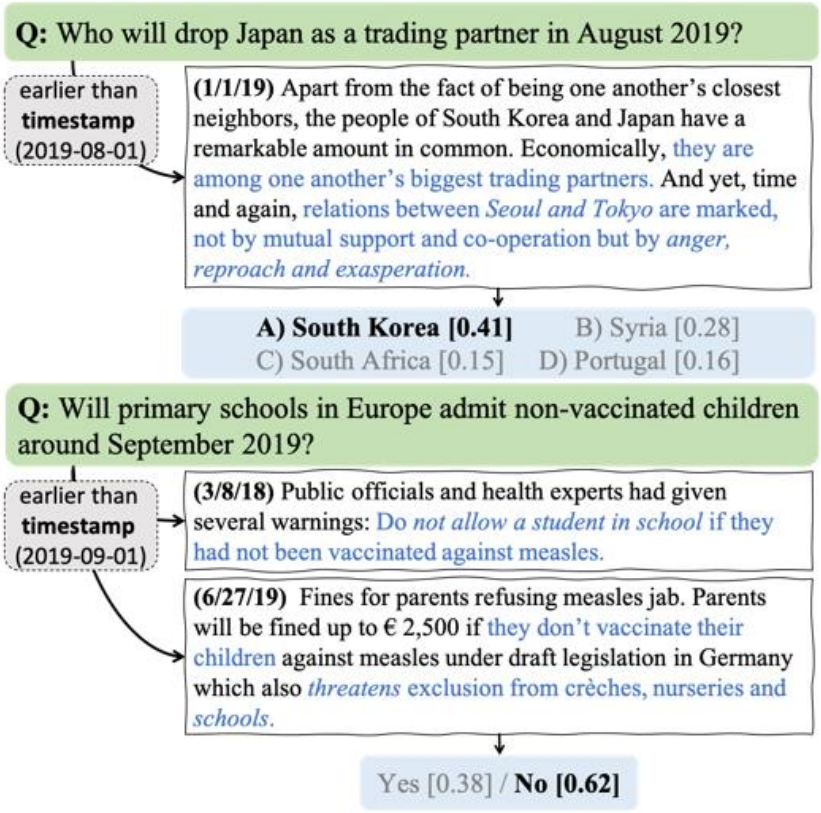
warm-up

Q4: What happened before a woman was trapped?
A: landslide
Q5: What had started before a woman was trapped?
A: snow rainfall landslide
Q6: What happened while a woman was trapped?
A: searching
Q7: What happened after a woman was trapped?
A: searching said found

User-provided

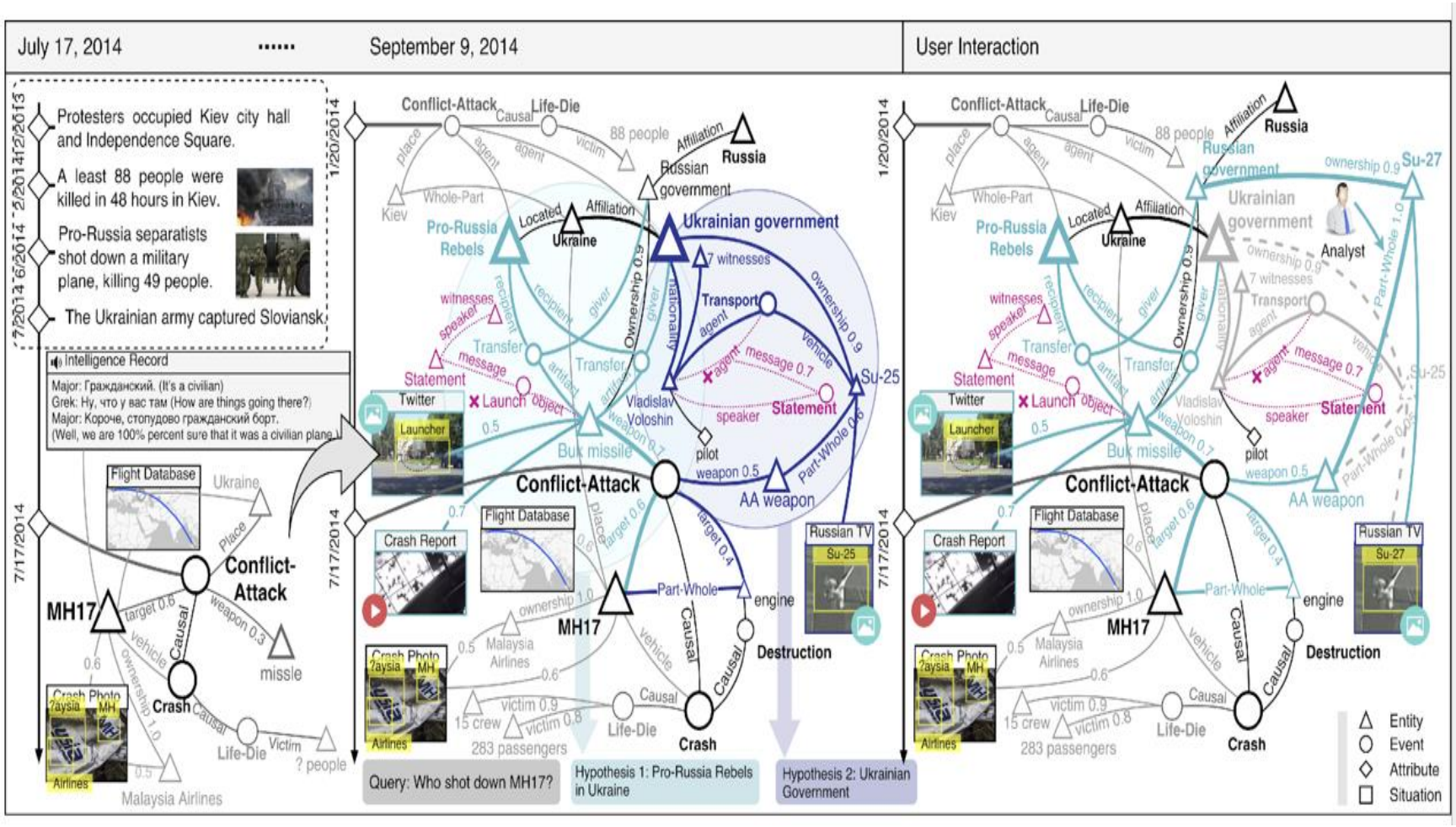
Ning, et al. TORQUE: A Reading Comprehension Dataset of Temporal Ordering Questions. EMNLP, 2020

- 3.2k news snippets with 21k human-generated questions querying temporal relationships



ForecastQA: A Question Answering Challenge for Event Forecasting

Intelligence Analysis



NLP in Industry

- Search (written and spoken)
- Online advertisement
- Automated/assisted translation

- Sentiment analysis for marketing or finance/trading
- Speech recognition
- Automating customer support

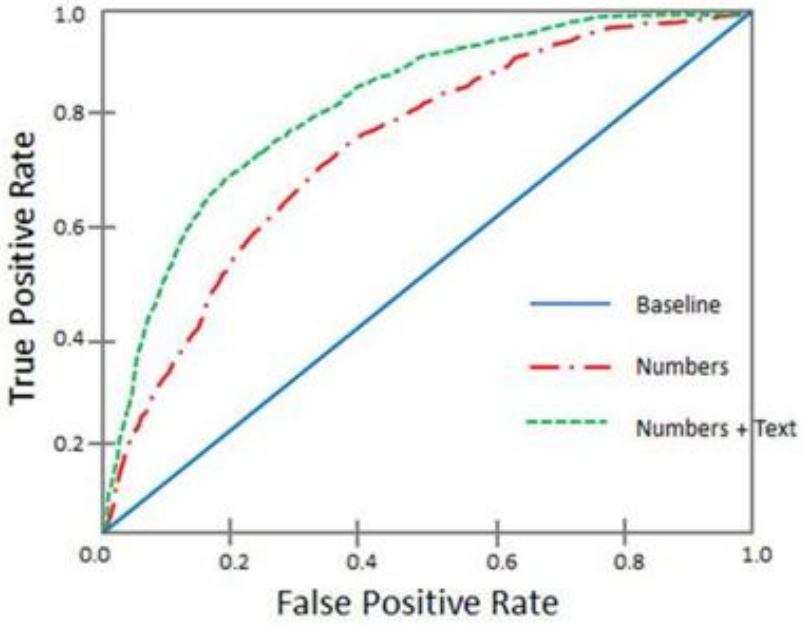
3/18/11 at 4:00 PM | 17 Comments

Mentions of the Name 'Anne Hathaway' May Drive Berkshire Hathaway Stock

By Patrick Huguenin

Structured + Text Data in Predictive Models

- Use of both types of data in building predictive models.



ROC Chart of Models With and Without Textual Comments

High-level NLP tasks

- Information Extraction
  - search, event detection, textual entailment
- Writing Assistance
  - spell checking, grammar checking, auto-completion
- Text Classification
  - spam, sentiment, author, plagiarism
- Natural language understanding
  - metaphor analysis, argumentation mining, question-answering
- Natural language generation
  - summarization, tutoring systems, chat bots
- Multilinguality
  - machine translation, cross-lingual information retrieval

Some Low-level NLP Tasks

- NLP applications require several NLP analyses:
  - Word tokenization
  - Sentence boundary detection
  - Part-of-speech (POS) tagging
    - to identify the part-of-speech (e.g. noun, verb) of each word
  - Named Entity (NE) recognition
    - to identify proper nouns (e.g. names of person, location, organization; domain terminologies)
  - Parsing
    - to identify the syntactic structure of a sentence
  - Semantic analysis
    - to derive the meaning of a sentence

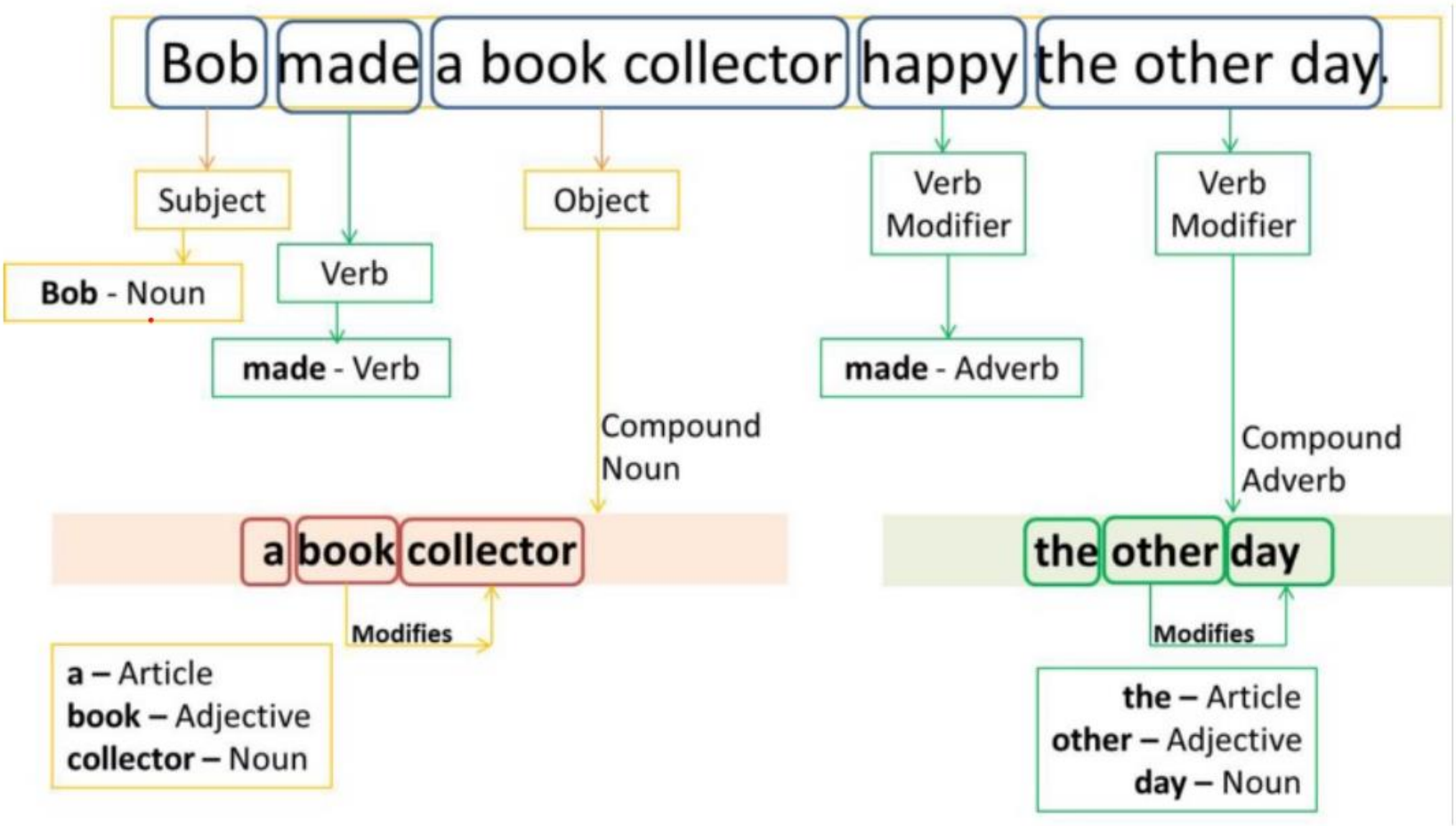
1. Part-Of-Speech (POS) Tagging

- POS tagging is a process of assigning a POS or lexical class marker to each word in a sentence (and all sentences in a corpus).

Input: `the lead paint is unsafe`

Output: `the/Det lead/N paint/N is/V unsafe/Adj`

Example


Bob made a book collector happy the other day.
Subject
Verb
Object
Verb Modifier
Verb Modifier
Bob - Noun
made - Verb
made - Adverb
Compound Noun
Compound Adverb
a book collector
the other day
Modifies
Modifies
a – Article
book – Adjective
collector – Noun
the – Article
other – Adjective
day – Noun

NLTK POS Tagger

POS tag list:

- CC coordinating conjunction
- CD cardinal digit
- DT determiner
- EX existential there (like: "there is" ... think of it like "there exists")
- FW foreign word
- IN preposition/subordinating conjunction
- JJ adjective 'big'
- JJR adjective, comparative 'bigger'
- JJS adjective, superlative 'biggest'
- LS list marker 1)
- MD modal could, will
- NN noun, singular 'desk'
- NNS noun plural 'desks'
- NNP proper noun, singular 'Harrison'
- NNPS proper noun, plural 'Americans'
- PDT predeterminer 'all the kids'
- POS possessive ending parent's
- •PRP personal pronoun I, he, she
- •PRP$ possessive pronoun my, his, hers
- •RB adverb very, silently,
- •RBR adverb, comparative better
- •RBS adverb, superlative best
- •RP particle give up
- •TO to go 'to' the store.
- •UH interjection errrrrrrrm
- •VB verb, base form take
- •VBD verb, past tense took
- •VBG verb, gerund/present participle taking
- •VBN verb, past participle taken
- •VBP verb, sing. present, non-3d take
- •VBZ verb, 3rd person sing. present takes
- •WDT wh-determiner which
- •WP wh-pronoun who, what
- •WP$ possessive wh-pronoun whose
- •WRB wh-abverb where, when

```
>>> text = word_tokenize("They refuse to permit us to obtain the refuse permit")
>>> nltk.pos_tag(text) [('They', 'PRP'), ('refuse', 'VBP'), ('to', 'TO'), ('permit', 'VB'), ('us', 'PRP'), ('to', 'TO'), ('obtain', 'VB'), ('the', 'DT'), ('refuse', 'NN'), ('permit', 'NN')]
```

Source : https://www.nltk.org/book/ch05.html

2. Named Entity Recognition (NER)

- NER is to process a text and identify named entities in a sentence
  - e.g. “U.N. official Ekeus heads for Baghdad.”

[ORG U.N.] official [PER Ekeus] heads for [LOC Baghdad] .

NER

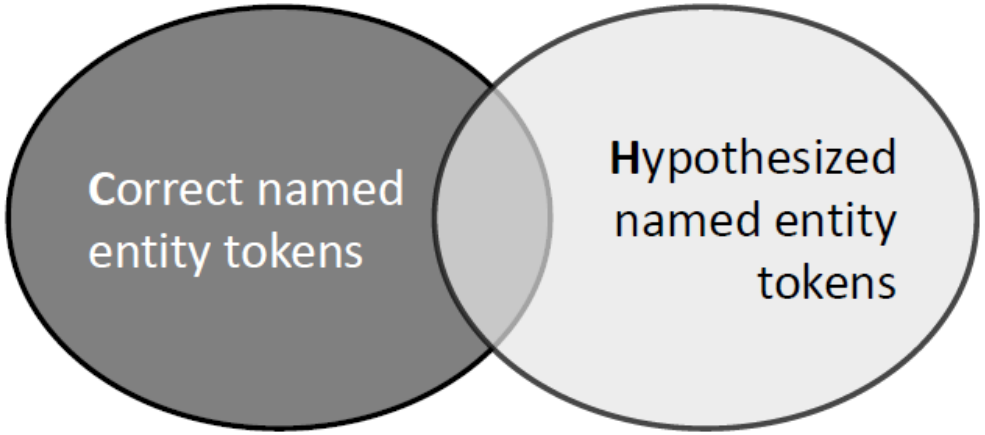
Elizabeth Warren, the liberal firebrand who emerged as a top Democratic contender for the **White House** on the strength of an anti-corruption platform backed by a dizzying array of policy proposals, ended her campaign on Thursday. A former bankruptcy law professor who forged a national reputation as a scourge of Wall Street even before entering politics, **Warren** had banked on a strong showing on Super Tuesday after a string of disappointing finishes in the early states. But she trailed far behind front-runners **Bernie Sanders** and **Joe Biden**, placing third in her home state of Massachusetts, which she continues to represent in the U.S. Senate.

- Label certain kinds of proper nouns:
  - Personal names
  - Organizations
  - Geopolitical entities
  - Locations
  - etc.

Some Named Entity Types

- Different annotation schemes for NER use different types
- Common types include
- PER—person
- ORG—Organization
- LOC—Location
- GPE—Geopolitical Entity
- FAC—Facility
- NAT—Natural phenomenon
- These are only tagged when they are proper names

Evaluating an NER System



$$\text{recall} = \frac{|C \cap H|}{|C|}$$

$$\text{precision} = \frac{|C \cap H|}{|H|}$$

NER System Building Process

NER System Building Process

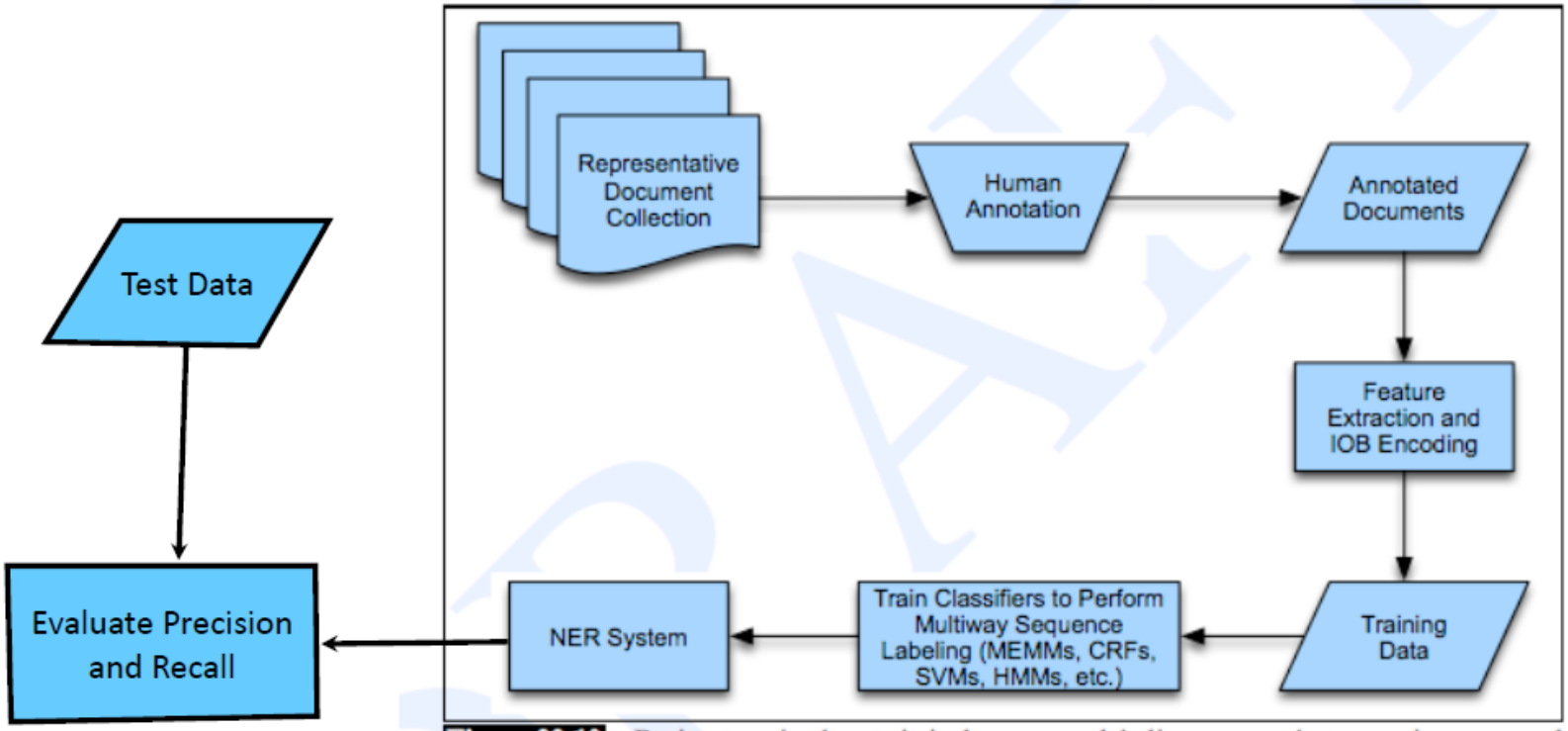


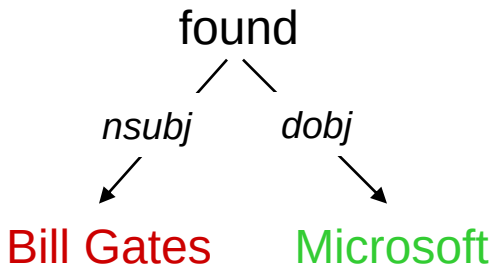
Figure 22.10 Basic steps in the statistical sequence labeling approach to creating a named entity recognition system.

3. Shallow Parsing

- **Shallow parsing** (also **chunking** or **light parsing**) is an analysis of a sentence which first identifies constituent parts of sentences (nouns, verbs, adjectives, etc.) and then links them to higher order units that have discrete grammatical meanings (noun groups or phrases, verb groups, etc.).
- Shallow (or Partial) parsing identifies the (base) syntactic phases in a sentence.

- After NEs are identified, **dependency parsing** is often applied to extract the syntactic/dependency relations between the NEs.

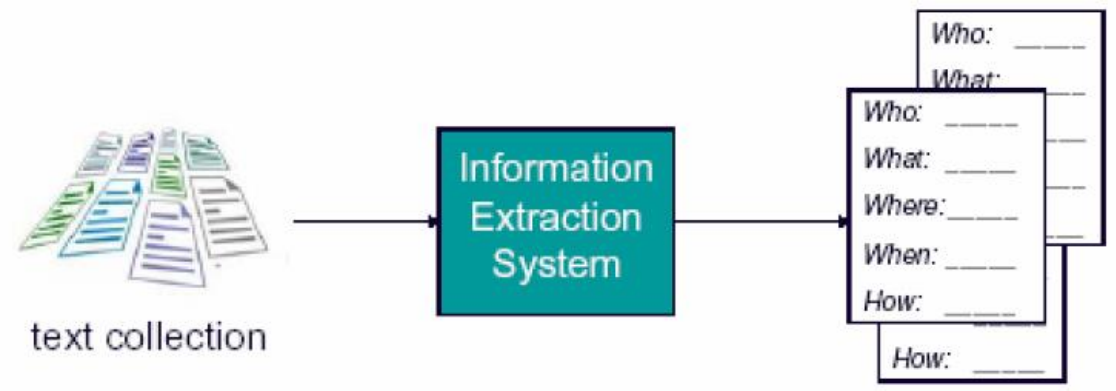
[PER Bill Gates] founded [ORG Microsoft].



Dependency Relations
nsubj(Bill Gates, found)
dobj(found, Microsoft)

4. Information Extraction (IE)

- Identify specific pieces of information (data) in an unstructured or semi-structured text
- Transform unstructured information in a corpus of texts or web pages into a structured database (or templates)
- Applied to various types of text, e.g.
  - Newspaper articles
  - Scientific articles
  - Web pages
  - etc.



Bridgestone Sports Co. said Friday it had set up a joint venture in Taiwan with a local concern and a Japanese trading house to produce golf clubs to be supplied to Japan.

The joint venture, Bridgestone Sports Taiwan Co., capitalized at 20 million new Taiwan dollars, will start production in January 1990 with production of 20,000 iron and "metal wood" clubs a month.

template filling

TIE-UP-1
Relationship: TIE-UP
Entities: "Bridgestone Sport Co."
"a local concern"
"a Japanese trading house"
Joint Venture Company:
"Bridgestone Sports Taiwan Co."
Activity: **ACTIVITY-1**
Amount: NT$200000000

ACTIVITY-1
Activity: PRODUCTION
Company:
"Bridgestone Sports Taiwan Co."
Product:
"iron and 'metal wood' clubs"
Start Date:
DURING: January 1990

5. Reference resolution

Elizabeth Warren, the liberal firebrand who emerged as a top Democratic contender for the **White House** on the strength of an anti-corruption platform backed by a dizzying array of policy proposals, ended her campaign on Thursday. A former bankruptcy law professor who forged a national reputation as a scourge of Wall Street even before entering politics, **Warren** had banked on a strong showing on Super Tuesday after a string of disappointing finishes in the early states. But she trailed far behind front-runners **Bernie Sanders** and **Joe Biden**, placing third in her home state of Massachusetts, which she continues to represent in the U.S. Senate.

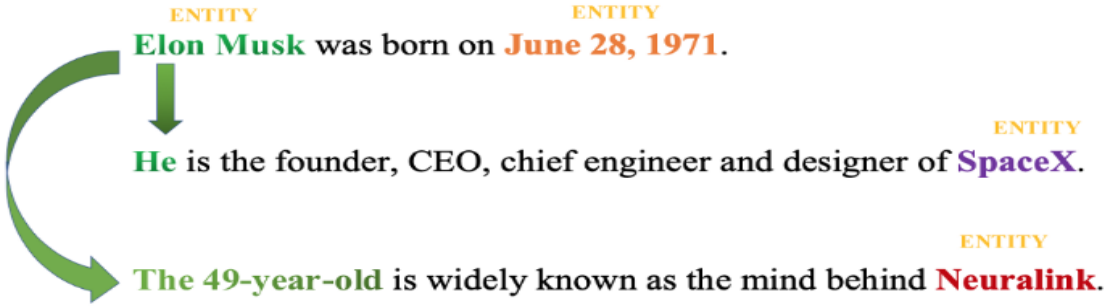
Reference, in NLP, is a linguistic process where one word in a sentence or discourse may refer to another word or entity. The task of resolving such references is known as ***Reference Resolution.***

Source : http://demo.clab.cs.cmu.edu/NLP/S21/files/slides/02-apps.pdf

6. Coreference resolution

Example Discourse:

"**Elon Musk** was born on June 28, 1971. **He** is the founder, CEO , chief engineer and designer of SpaceX. **The 49 year old** is widely known as the mind behind Neuralink."



Referring Expressions: Elon Musk, He, The 49 year old

Referent: *Elon Musk*

Corefering Expressions: {Elon Musk, He}, {Elon Musk, The 49 year old}

The task of **coreference resolution** makes use of knowledge about how words like *that* or pronouns like *it* or *she* refer to previous parts of the **discourse**.

Coreference Resolution in particular, is the process of resolving pronouns to identify which entities are they referring to. It is also a kind of Reference Resolution. The entities resolved may be a person, place, organization, or event.

Discourse in the context of NLP refers to a sequence of sentences occurring one after the other. There will obviously be entities that are being talked about and possible references to those entities in the discourse.

Source: https://towardsai.net/p/nlp/c-r#:~:text=Reference%2C%20in%20NLP%2C%20is%20a,is%20known%20as%20Reference%20Resolution.

7. Relation detection/extraction

WASHINGTON/SELMA, Ala. (Reuters) - Democratic U.S. presidential front-runner Bernie Sanders raised $46.5 million in February, his campaign said on Sunday, and will launch new television ad buys in nine states with primaries later this month after this week's Super Tuesday contests. Joe Biden's campaign reported raising $5 million the day of the South Carolina primary. His February haul was $18 million, spokesman Michael Gwin said. Meanwhile, rival Elizabeth Warren, who struggled to a fifth-place finish in South Carolina, raised more than $29 million in February, her campaign manager Roger Lau said in a memo to supporters on Sunday.

Candidate	member_of
Bernie Sanders	Democratic Party
Joe Biden	Democratic Party
Elizabeth Warren	Democratic Party

- Relationship detection/extraction is the task of extracting semantic relationships from a text.
- Extracted relationships usually occur between two or more entities of
  - a certain type (e.g. Person, Organisation, Location) and fall into
  - a number of semantic categories (e.g. married to, employed by, lives in).

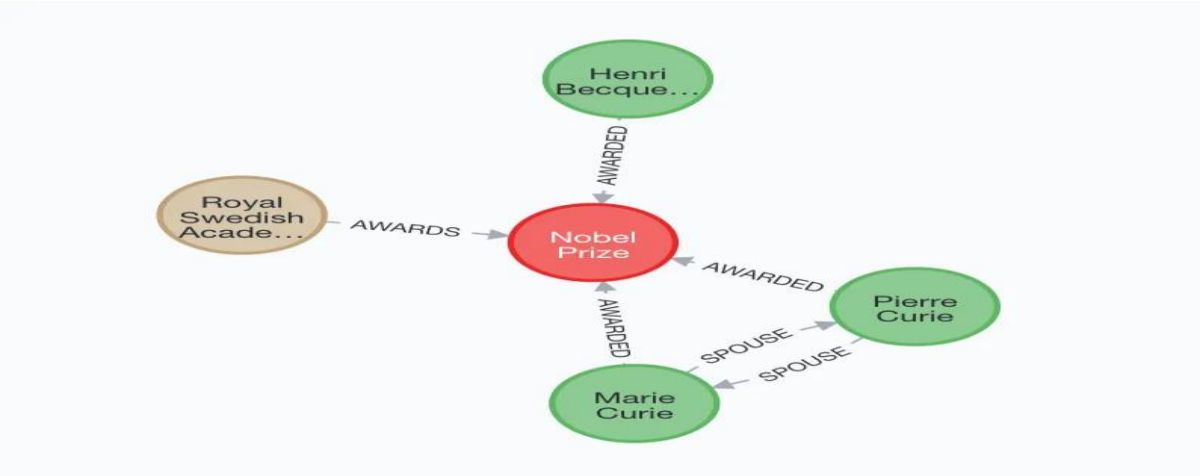
Examples of Relations

Relations		Examples	Types
Affiliations			
	Personal	*married to, mother of*	PER → PER
	Organizational	*spokesman for, president of*	PER → ORG
	Artifactual	*owns, invented, produces*	(PER | ORG) → ART
Geospatial			
	Proximity	*near, on outskirts*	LOC → LOC
	Directional	*southeast of*	LOC → LOC
Part-Of			
	Organizational	*a unit of, parent of*	ORG → ORG
	Political	*annexed, acquired*	GPE → GPE

Figure 22.11 Semantic relations with examples and the named entity types they involve.

Ex: Named Entity Recognition and Relationship Extraction

- **Building Named Entity Recognition and Relationship Extraction Components with HuggingFace Transformers (https://odsc.medium.com/building-named-entity-recognition-and-relationship-extraction-components-with-huggingface-77d233e27e65)**



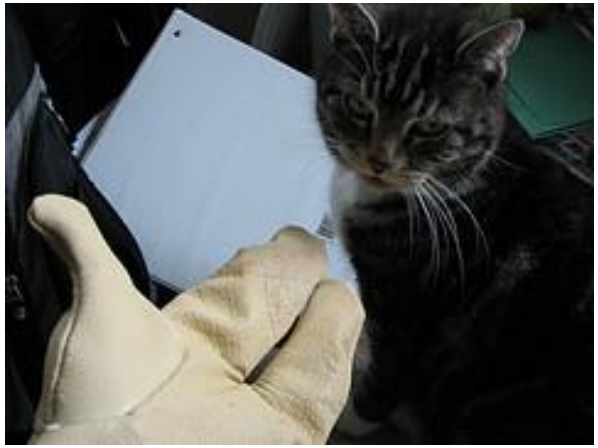
Relations detected by a Relation Extraction (RE) component

But NLP very is hard..

- Understanding natural languages is hard … because of inherent *ambiguity*
- Engineering NLP systems is also hard … because of:
  - Huge amount of data resources needed (e.g. grammar, dictionary, documents to extract statistics from)
  - Computational complexity (intractable) of analyzing a sentence

Ambiguity (1)

“Get the cat with the gloves.”



Source: Marti Hearst, i256, at UC Berkeley

Ambiguity (2)

Find at least 5 meanings of this sentence:

"I made her duck"

1. I cooked waterfowl for her benefit (to eat)
2. I cooked waterfowl belonging to her
3. I created the (plaster?) duck she owns
4. I caused her to quickly lower her head or body
5. I waved my magic wand and turned her into undifferentiated waterfowl

Ambiguity (3)

Some ambiguous headlines

- Juvenile Court to Try Shooting Defendant
- Teacher Strikes Idle Kids
- Kids Make Nutritious Snacks
- Bush Wins on Budget, but More Lies Ahead
- Hospitals are Sued by 7 Foot Doctors

Source: Marti Hearst, i256, at UC Berkeley

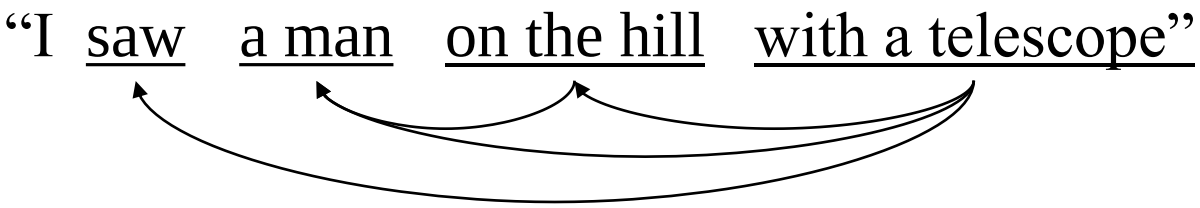
Ambiguity is Pervasive

- **Phonetics**
  - I mate or duck
  - I'm eight or duck
  - Eye maid; her duck
  - Aye mate, her duck
  - I maid her duck
  - I'm aid her duck
  - I mate her duck
  - I'm ate her duck
  - I'm ate or duck
  - I mate or duck

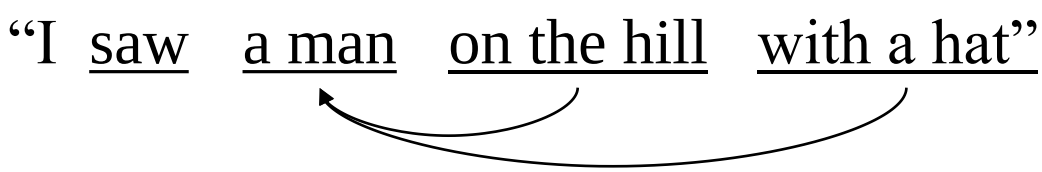
Sound like
"I made her duck"

Source: Jurafsky & Martin "Speech and Language Processing"

- **Lexical category** (part-of-speech)
  - "duck" as a noun or a verb
- **Lexical Semantics** (word meaning)
  - "duck" as an animal or a plaster duck statue
- Compound nouns
  - e.g. "dog food", "Intelligent design scores …"
- **Syntactic ambiguity**



- [But semantics can sometimes help disambiguate]



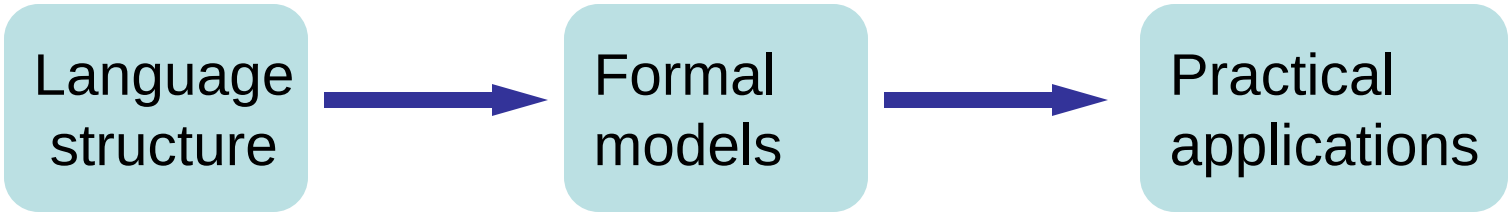
The Bottom Line

- Complete NL Understanding (thus general intelligence) is impossible.
- But we can make incremental progress.
- Also we have made successes in **limited domains**.

The Big Picture Approach

All of these applications operate by **exploiting underlying regularities** in human languages.

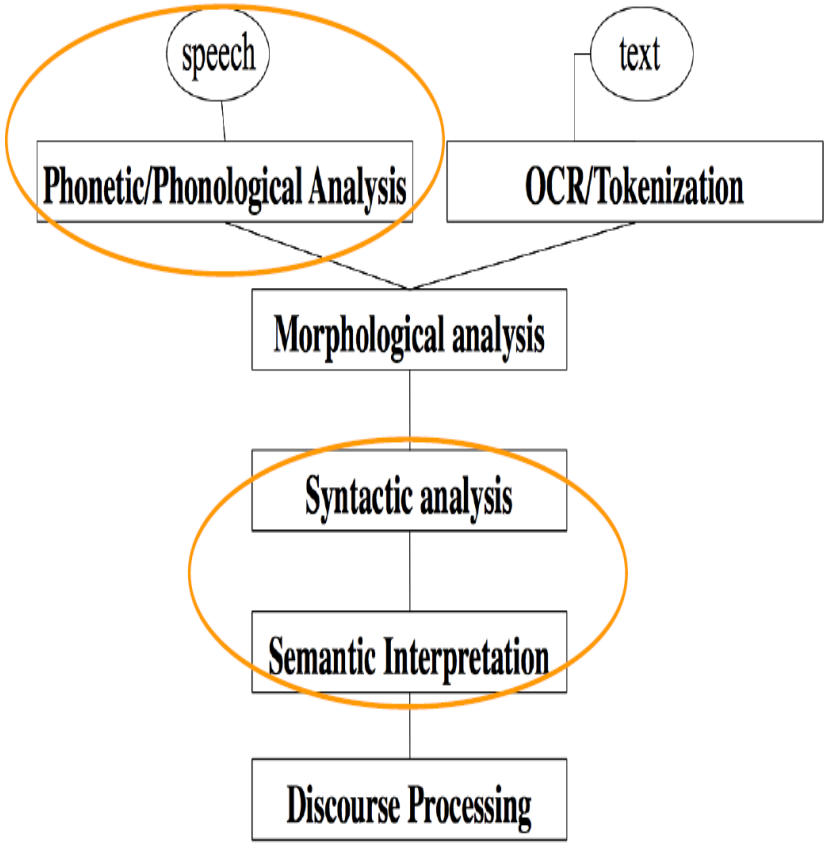
Sometimes in complex ways, sometimes in pretty trivial ways.



Topics: Linguistics

- Word-level processing
- Syntactic processing
- Lexical and compositional semantics
- Discourse structure

NLP Levels



NLP requires various kinds of knowledge of language:

- Phonetics and Phonology —knowledge about linguistic sounds
- Morphology —knowledge of the meaningful components of words
- Syntax —knowledge of the structural relationships between words
- Semantics —knowledge of meaning
- Pragmatics — knowledge of the relationship of meaning to the goals and intentions of the speaker
- Discourse —knowledge about linguistic units larger than a single utterance

Different Levels of Linguistic Analysis

- Phonology
  - Speech audio signal to phonemes
- Morphology
  - Inflection (e.g. “I”, “my”, “me”; “eat”, “eats”, “ate”, “eaten”)
  - Derivation (e.g. “teach”, “teacher”, “nominate”, “nominee”)
- Syntax
  - Part-of-speech (noun, verb, adjective, preposition, etc.)
  - Phrase structure (e.g. noun phrase, verb phrase)
- Semantics
  - Meaning of a word (e.g. “book” as a bound volume or an accounting ledger) or a sentence
- Discourse
  - Meaning and inter-relation between sentences

Topics: Techniques

- Finite-state methods
- Context-free methods
- Probabilistic models
- Neural network models

Supervised machine learning methods

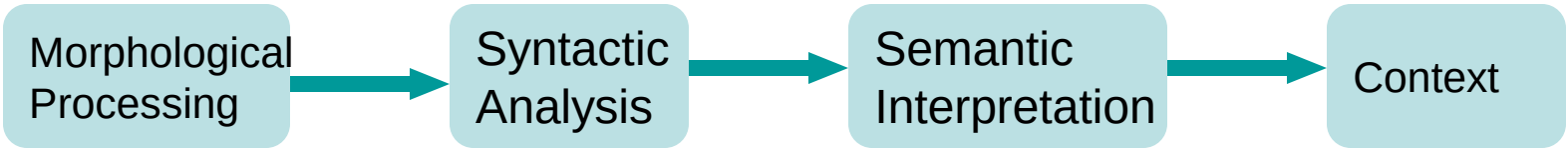
Process Pipeline

- Phonology
- Morphology
- Syntax
- Semantics
- Pragmatics
- Discourse

Each kind of knowledge has associated with it an encapsulated set of processes that make use of it.

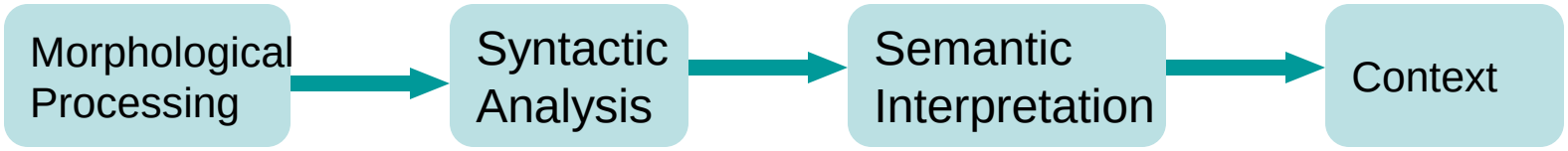
Interfaces are defined that allow the various levels to communicate.

This often leads to a **pipeline architecture**.

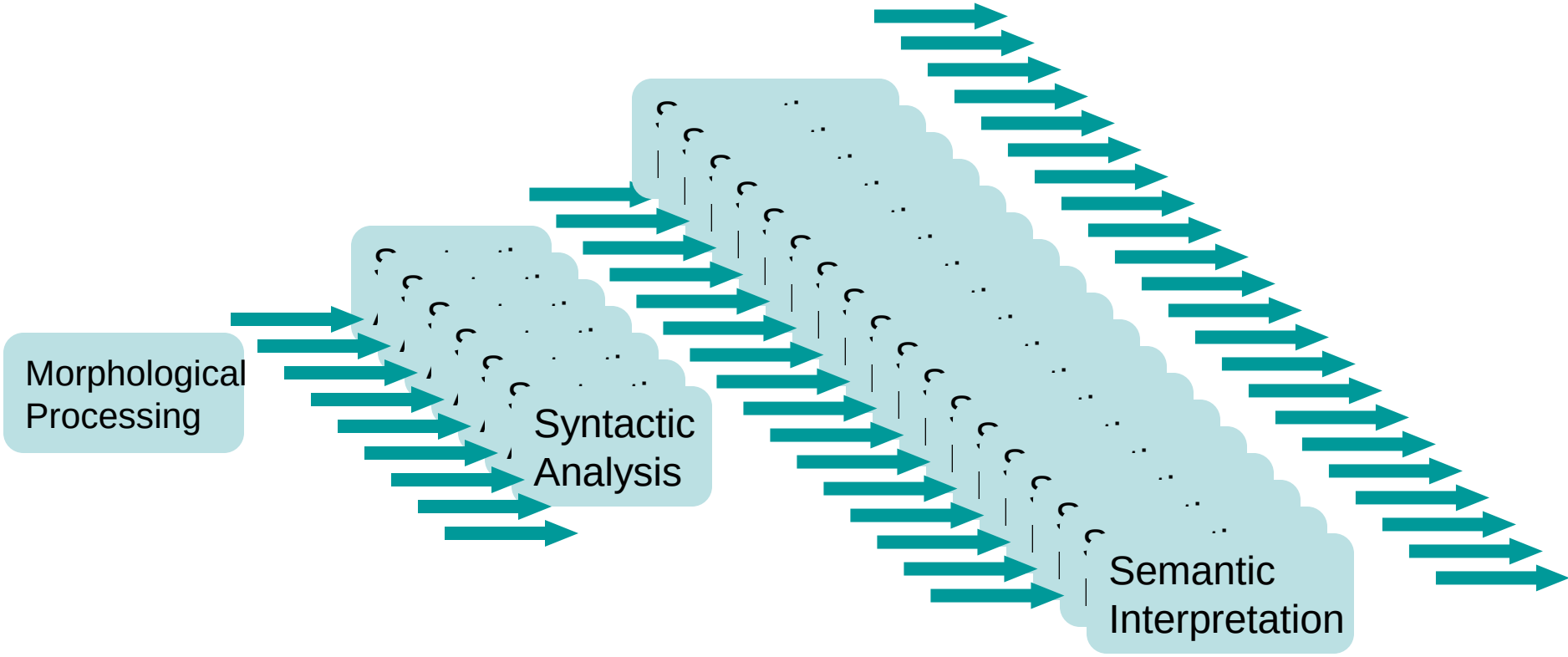


But Problem..

- Remember our pipeline...



It really looks like this



Dealing with Ambiguity

Four possible approaches:

1. **Formal approaches** -- Tightly coupled interaction among processing levels; knowledge from other levels can help decide among choices at ambiguous levels.
2. Pipeline processing that ignores ambiguity as it occurs and hopes that other levels can eliminate incorrect structures.
3. **Probabilistic approaches** based on making the most likely choices
4. Don't do anything, maybe it won't matter

Models and Algorithms

- By models we mean the formalisms that are used to capture the various kinds of linguistic knowledge we need.
- Algorithms are then used to manipulate the knowledge representations needed to tackle the task at hand.

Various Models

- Finite state machines
- Rule-based and logic-based approaches
- Probabilistic models
- Neural network models

Various Algorithms

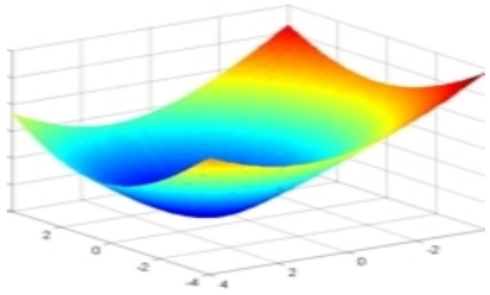
- In particular..
  - State-space search
    - To manage the problem of making choices during processing when we lack the information needed to make the right choice
  - Dynamic programming
    - To avoid having to redo work during the course of a state-space search
      - CKY, Earley, Minimum Edit Distance, Viterbi, Baum-Welch
  - Classifiers
    - Machine learning based classifiers that are trained to make decisions based on features extracted from the local context

What's Deep Learning (DL)?

- Deep learning is a subfield of machine learning
- Most machine learning methods work well because of human-designed representations and input features
  - For example: features for finding named entities like locations or organization names (Finkel, 2010):

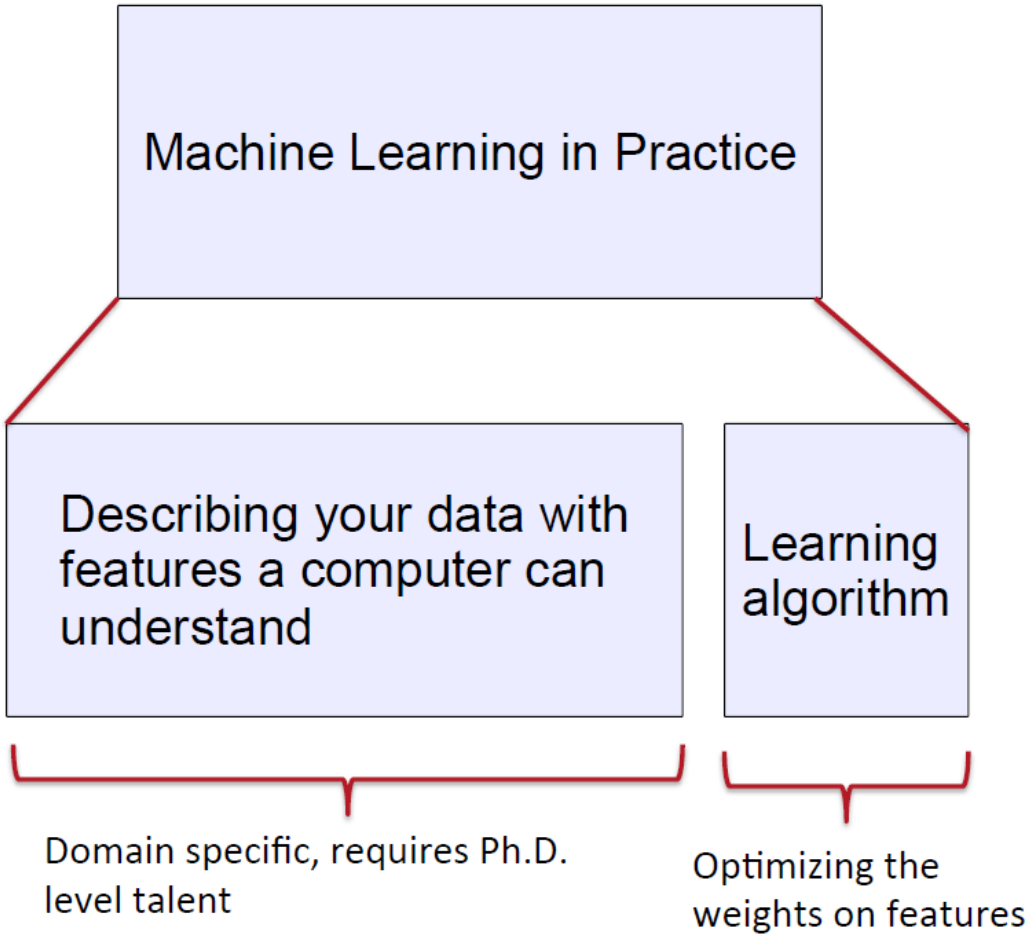
Feature	NER
Current Word	✓
Previous Word	✓
Next Word	✓
Current Word Character n-gram	all
Current POS Tag	✓
Surrounding POS Tag Sequence	✓
Current Word Shape	✓
Surrounding Word Shape Sequence	✓
Presence of Word in Left Window	size 4
Presence of Word in Right Window	size 4

- Machine learning becomes just optimizing weights to best make a final prediction



Source : cs224d

Machine Learning vs Deep Learning



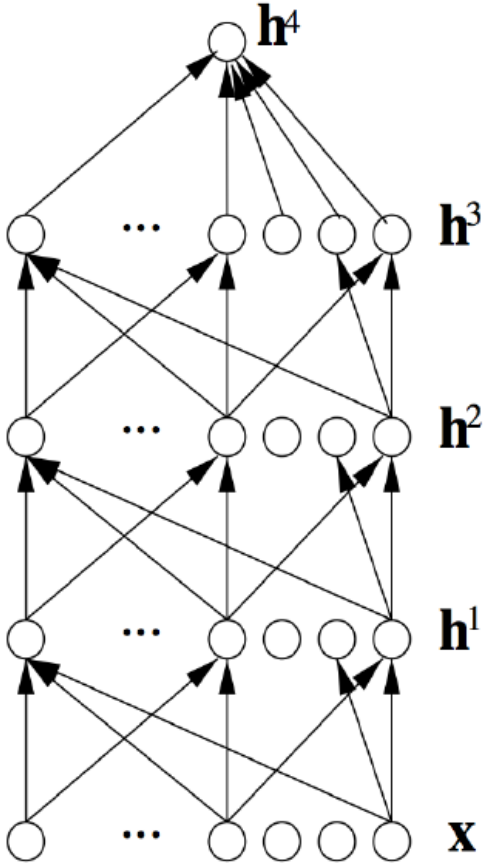
Source : cs224d

What's Deep Learning (DL)?

- Representation learning attempts to automatically learn good features or representations

- Deep learning algorithms attempt to learn (multiple levels of) representation and an output

- From "raw" inputs **x** (e.g. words)



Source : cs224d

Reasons for Exploring Deep Learning

- Manually designed features are often over-specified, incomplete and take a long time to design and validate
- **Learned Features** are easy to adapt, fast to learn

- Deep learning provides a very flexible, (almost?) universal, learnable framework for **representing** world, visual and linguistic information.

- Deep learning can learn **unsupervised** (from raw text) and **supervised** (with specific labels like positive/negative)

Source : cs224d

Reasons for Exploring Deep Learning

- In 2006 **deep** learning techniques started outperforming other machine learning techniques. Why now?

- DL techniques benefit more from a lot of data
- Faster machines and multicore CPU/GPU help DL
- New models, algorithms, ideas

→ **Improved performance** (first in speech and vision, then NLP)

Deep Learning + NLP = Deep NLP

- Combine ideas and goals of NLP and use representation learning and deep learning methods to solve them

- Several big improvements in recent years across different NLP
  - **levels**: speech, morphology, syntax, semantics
  - **applications**: machine translation, sentiment analysis and question answering

Source : cs224d