Classification Applications

CS4248 Natural Language Processing

Week 11

Min-Yen KAN

11

Recap of Week 10

Lexical Semantics

- Meaning of and relationships among words
- WordNet 3.0 hand-annotated many of these relationships
- Compute relatedness via ontological tree structure

Logical Semantics

- Meaning representation with FOL with terms and relations
- Unify (fill expected arguments appropriately) while parsing

Week 11 Agenda

Sentiment Analysis

Summarization I

Question Answering I

There are many more, but these are common sample tasks

What is Sentiment Analysis?

A.k.a. Opinion extraction, Opinion mining, Sentiment mining, Subjectivity analysis

Slide Credits: Dan Jurafsky (Stanford)

Google Product Search

HP Officejet 6500A Plus e-All-in-One Color Ink-jet - Fax / copier / printer / scanner
$89 online, **$100** nearby ★★★★☆ 377 reviews
September 2010 - Printer - HP - Inkjet - Office - Copier - Color - Scanner - Fax - 250 sh

Reviews

Summary - Based on 377 reviews

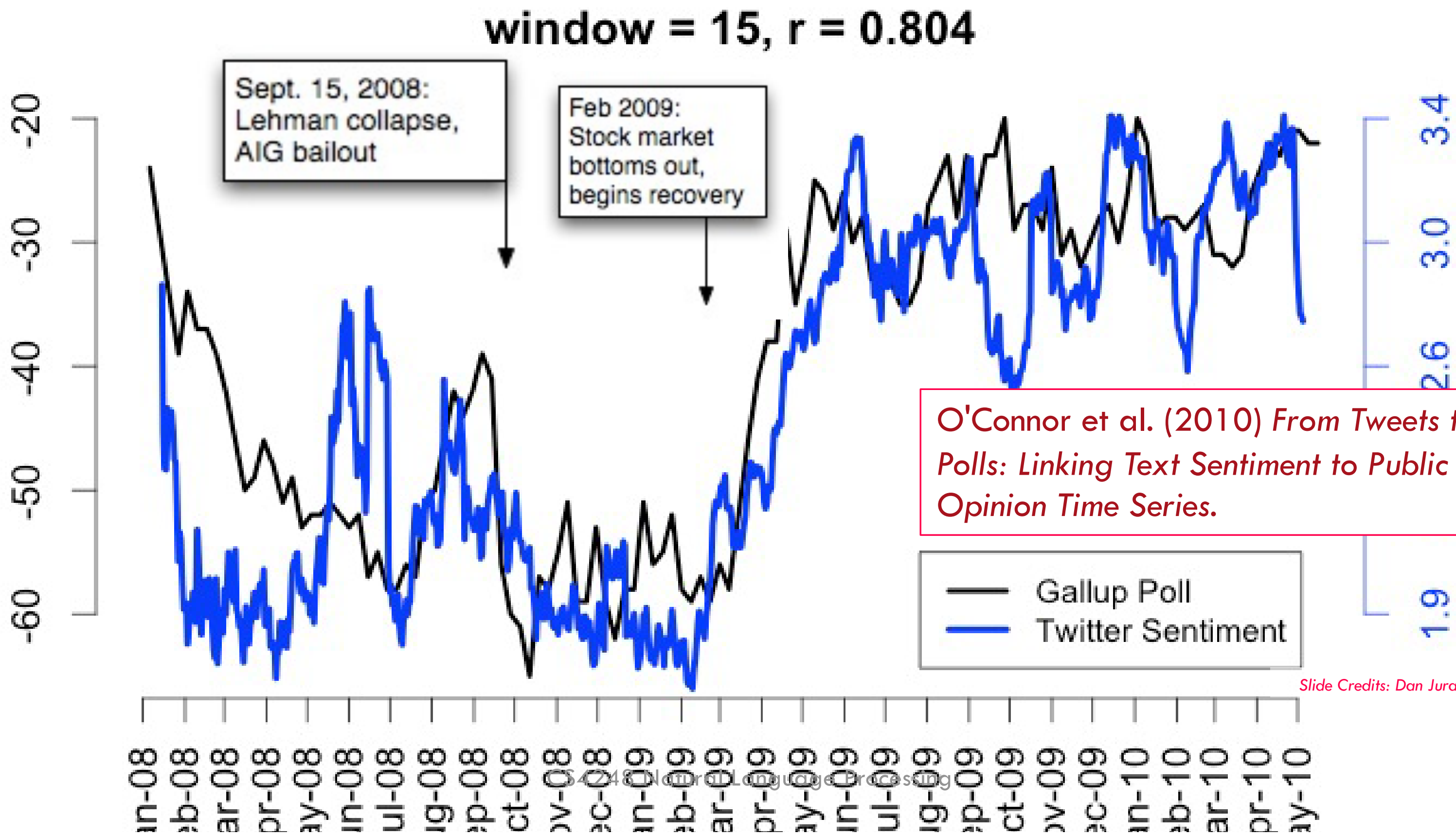
1 star | 2 | 3 | 4 stars | 5 stars

What people are saying

ease of use	"This was very easy to setup to four computers."
value	"Appreciate good quality at a fair price."
setup	"Overall pretty easy setup."
customer service	"I DO like honest tech support people."
size	"Pretty Paper weight."
mode	"Photos were fair on the high quality mode."
colors	"Full color prints came out with great quality."

Slide Credits: Dan Jurafsky (Stanford)

Twitter sentiment versus Gallup Poll of Consumer Confidence

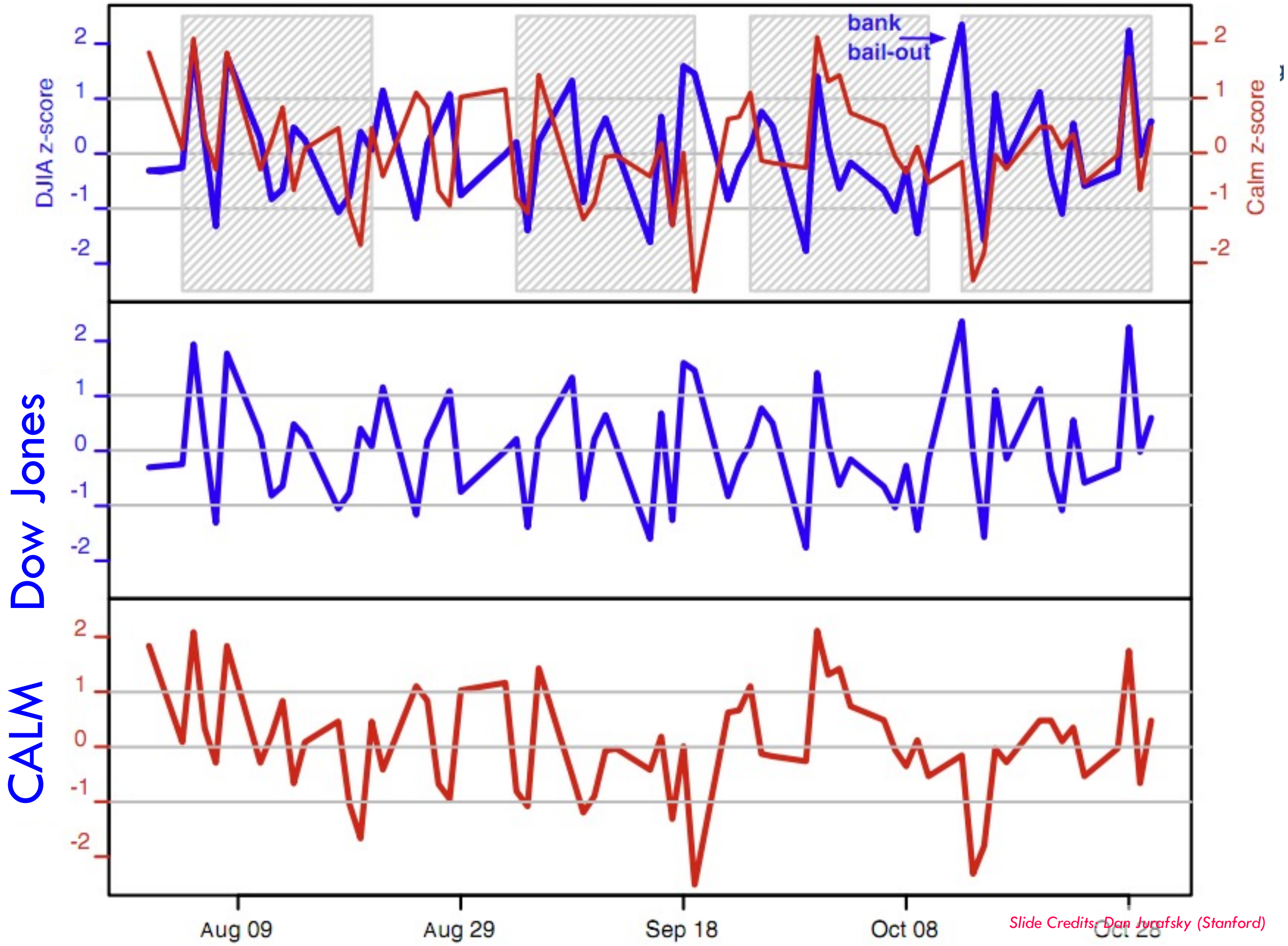


O'Connor et al. (2010) *From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series.*

Slide Credits: Dan Jurafsky (Stanford)

- CALM predicts DJIA 3 days later
- At least one current hedge fund uses this algorithm

Bollen et al. (2011) *[Twitter mood predicts the stock market.](#)*



Slide Credits: Dan Jurafsky (Stanford)

Why sentiment analysis?

- Movie: is this review positive or negative?
- Products: what do people think about the new iPhone?
- Public sentiment: how is consumer confidence? Is despair increasing?
- Politics: what do people think about this candidate or issue?
- Prediction: predict election outcomes or market trends from sentiment

Slide Credits: Dan Jurafsky (Stanford)

Scherer Typology of Affective States

- Emotion: brief organically synchronized … evaluation of a major event
  - angry, sad, joyful, fearful, ashamed, proud, elated
- Mood: diffuse non-caused low-intensity long-duration change in subjective feeling
  - cheerful, gloomy, irritable, listless, depressed, buoyant
- Interpersonal stances: affective stance toward another person in a specific interaction
  - friendly, flirtatious, distant, cold, warm, supportive, contemptuous
- Attitudes: enduring, affectively colored beliefs, dispositions towards objects or persons
  - liking, loving, hating, valuing, desiring
- Personality traits: stable personality dispositions and typical behavior tendencies
  - nervous, anxious, reckless, morose, hostile, jealous

Slide Credits: Dan Jurafsky (Stanford)

Sentiment Analysis

Sentiment analysis is the detection of **attitudes**:
"enduring, affectively colored beliefs, dispositions towards objects or persons"

- Holder (source) of attitude
- Target (aspect) of attitude
- Type of attitude
  - From an enumerated set of types
    - Like, love, hate, value, desire, etc.
  - Or (more commonly) simple weighted polarity:
    - positive, negative, neutral, together with strength
- Text containing the attitude
  - Sentence or entire document

Slide Credits: Dan Jurafsky (Stanford)

Sentiment Analysis

Simplest task:

- Is the attitude of this text positive or negative?

More complex:

- Rank the attitude of this text from 1 to 5

Advanced:

- Detect the target, source, or complex attitude types

Slide Credits: Dan Jurafsky (Stanford)

Sentiment Baselines

IMDB data in the Pang and Lee database

✓

when _star wars_ came out some twenty years ago , the image of traveling throughout the stars has become a commonplace image . […]

when han solo goes light speed , the stars change to bright lines , going towards the viewer in lines that converge at an invisible point

cool .

october sky offers a much simpler image–that of a single white dot , traveling horizontally across the night sky . [. . .]

✗

“ snake eyes ” is the most aggravating kind of movie : the kind that shows so much potential then becomes unbelievably disappointing .

it's not just because this is a brian depalma film , and since he's a great director and one who's films are always greeted with at least some fanfare .

and it's not even because this was a film starring nicolas cage and since he gives a brauvara performance , this film is hardly worth his talents .

Slide Credits: Dan Jurafsky (Stanford)

Baseline Algorithm

(adapted from Pang and Lee)

1. Tokenization
2. Feature Extraction
3. Supervised classification using different classifiers
   - Naïve Bayes
   - Maximum Entropy
   - Support Vector Machine

Slide Credits: Dan Jurafsky (Stanford)

Sentiment Tokenization Issues

- Deal with HTML and XML markup
- Twitter mark-up (names, hash tags)
- Capitalization (preserve for words in all caps)
- Phone numbers, dates
- Emoticons

Potts emoticons

```
[<>]?                        # optional hat/brow
[:;=8]                       # eyes
[\-o\*\']?                   # optional nose
[\)\]\(\[dDpP/\:\}\{@\|\\]   # mouth
|                            #### reverse orientation
[\)\]\(\[dDpP/\:\}\{@\|\\]   # mouth
[\-o\*\']?                   # optional nose
[:;=8]                       # eyes
[<>]?                        # optional hat/brow
```

Slide Credits: Dan Jurafsky (Stanford)

Negation

Add NOT_ to every word between negation and following punctuation:

didn't like this movie , but I

didn't NOT_like NOT_this NOT_movie but I

Das and Chen (2001) *Yahoo! for Amazon: Extracting market sentiment from stock message boards.*
Pang et al. (2002) *Thumbs up? Sentiment Classification using Machine Learning Techniques.*

Slide Credits: Dan Jurafsky (Stanford)

Binarized Naïve Bayes

For sentiment (and some text classification tasks),
word occurrence may matter more than word frequency

- The occurrence of the word *fantastic* tells us a lot
- The fact that it occurs 5 times may not tell us much more.

Use Boolean Multinomial Naïve Bayes

- Clips all the word counts in each document to 1

Slide Credits: Dan Jurafsky (Stanford)

Subtlety and sarcasm in reviews

Perfume review in Perfumes: the Guide:

> *If you are reading this because it is your darling fragrance, …*
> *please wear it at home exclusively, and tape the windows shut.*

Dorothy Parker on Katherine Hepburn

> *She runs the gamut of emotions from A to B*

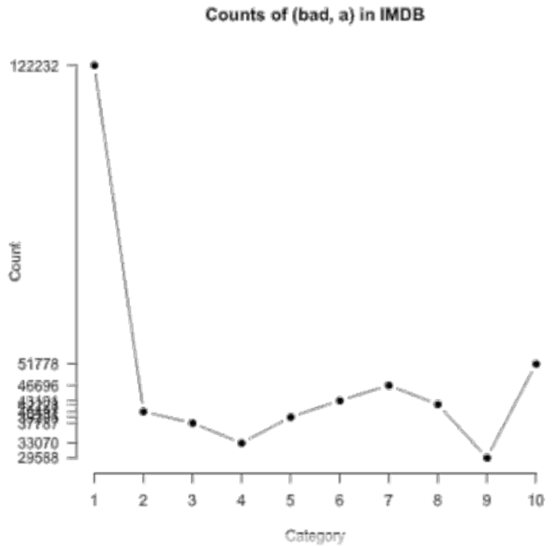
Slide Credits: Dan Jurafsky (Stanford)

Analyzing the polarity of each word in IMDB

How likely is each word to appear in each sentiment class?

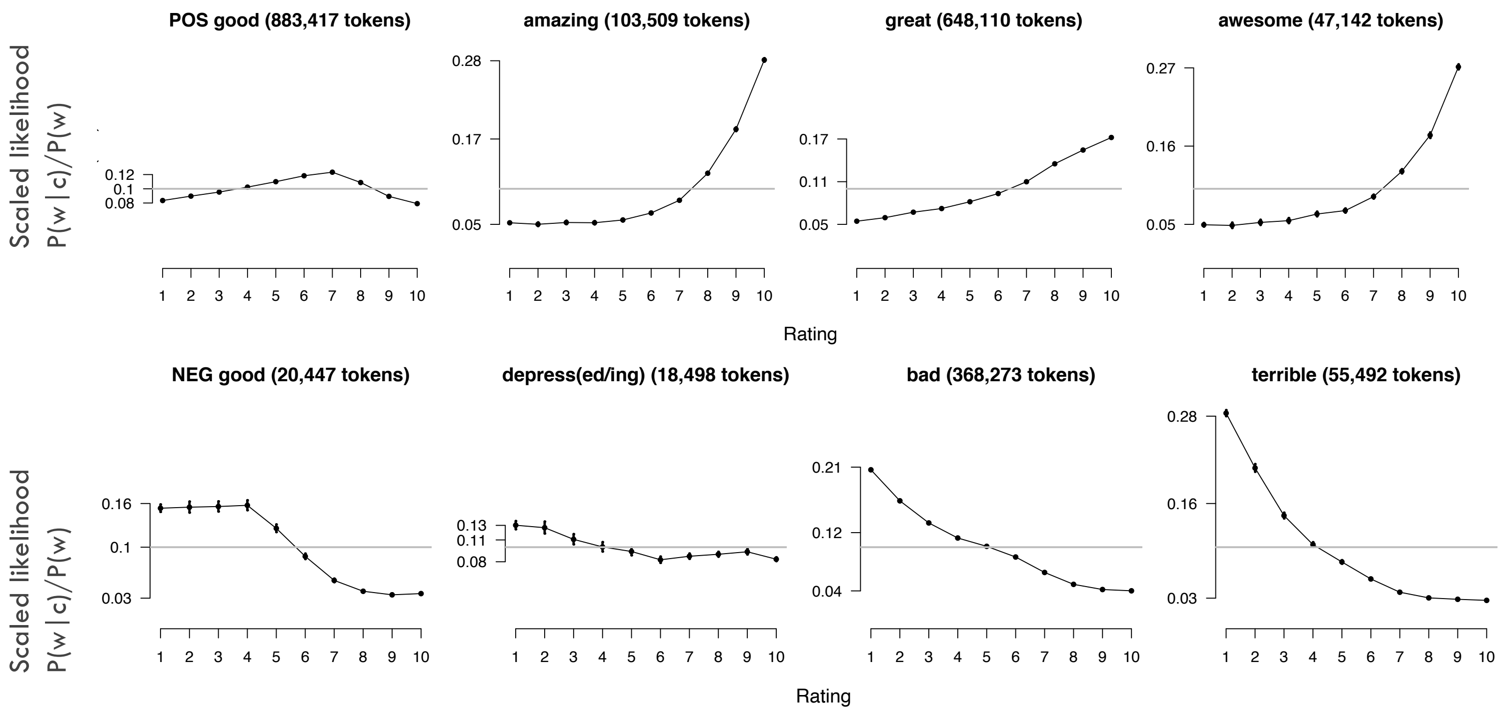
Potts (2011) *On the negativity of negation.*

- Count("*bad*") in 1-star, 2-star, 3-star, etc.
- Use likelihoods normalized from raw counts: $P(w|c) = \frac{f(w,c)}{\sum_{w \in c} f(w,c)}$
- Further use scaled likelihood to make them comparable among words: $\frac{P(w|c)}{P(w)}$



Slide Credits: Dan Jurafsky (Stanford)

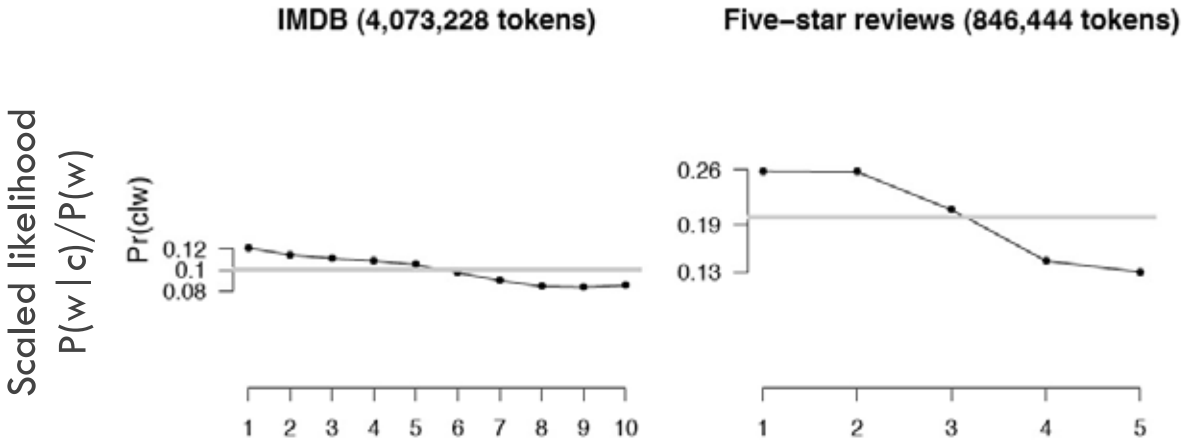
Analyzing the polarity of each word in IMDB



Slide Credits: Dan Jurafsky (Stanford)

More negation in negative sentiment

Do the same with logical negation (*not*, *n't*, *no*, *never*)



Slide Credits: Dan Jurafsky (Stanford)

Ordering Effects

This film should be brilliant. It sounds like a great plot, the actors are first grade, and the supporting cast is good as well, and Stallone is attempting to deliver a good performance. However, it can't hold up.

Well as usual Keanu Reeves is nothing special, but surprisingly, the very talented Laurence Fishbourne is not so good either, I was surprised.

Slide Credits: Dan Jurafsky (Stanford)

(Learning) Sentiment Lexicons

Many Sentiment Lexica

- The General Inquirer: http://www.wjh.harvard.edu/~inquirer
- LIWC (Linguistic Inquiry and Word Count) http://www.liwc.net/
- MPQA Subjectivity Cues Lexicon http://www.cs.pitt.edu/mpqa/subj_lexicon.html
- Bing Liu Opinion Lexicon http://www.cs.uic.edu/~liub/FBS/opinion-lexicon-English.rar
- SentiWordNet http://sentiwordnet.isti.cnr.it/

Slide Credits: Dan Jurafsky (Stanford)

Disagreements between polarity lexicons

	Opinion Lexicon	General Inquirer	SentiWordNet	LIWC
MPQA	33/5402 (**0.6%**)	49/2867 (**2%**)	1127/4214 (**27%**)	12/363 (**3%**)
Opinion Lexicon		32/2411 (**1%**)	1004/3994 (**25%**)	9/403 (**2%**)
General Inquirer			520/2306 (**23%**)	1/204 (**0.5%**)
SentiWordNet				174/694 (**25%**)

Christopher Potts (2011) *Sentiment Tutorial*

Slide Credits: Dan Jurafsky (Stanford)

Semi-supervised learning of lexicons

Grow a lexicon from seed examples

- Use a small amount of manual work
  - A few labeled examples
  - A few hand-built patterns
- Iterate, to build a lexicon

- Issues with noise creep

Slide Credits: Dan Jurafsky (Stanford)

Intuition for propagating word polarity

Insight: Adjectives conjoined by "*and*" have same polarity

Fair and legitimate, corrupt and brutal

? *fair and brutal*, ? *corrupt and legitimate*

Adjectives conjoined by "*but*" do not

fair but brutal

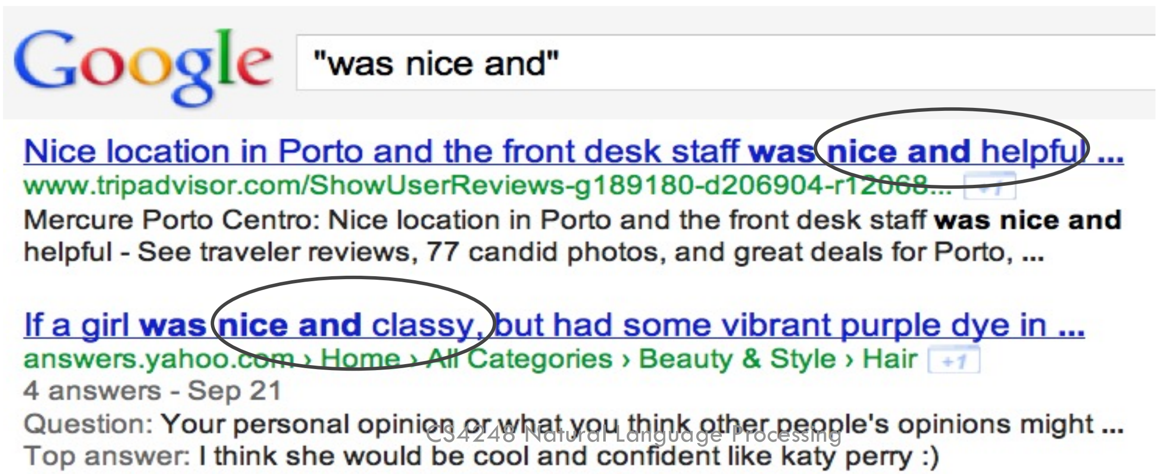
Hatzivassiloglou and McKeown (1997) *Predicting the Semantic Orientation of Adjectives.*

Slide Credits: Dan Jurafsky (Stanford)

"Snowball" algorithm

1. Label a seed set of adjectives
   +ve: *adequate helpful clever nice* …
   –ve: *contagious corrupt ignorant* …
2. Expand the seed set to conjoined adjectives via corpus search
   {grep|google} "*was nice and*"

We can learn even taxonomic hyper/hyponyms also from text!



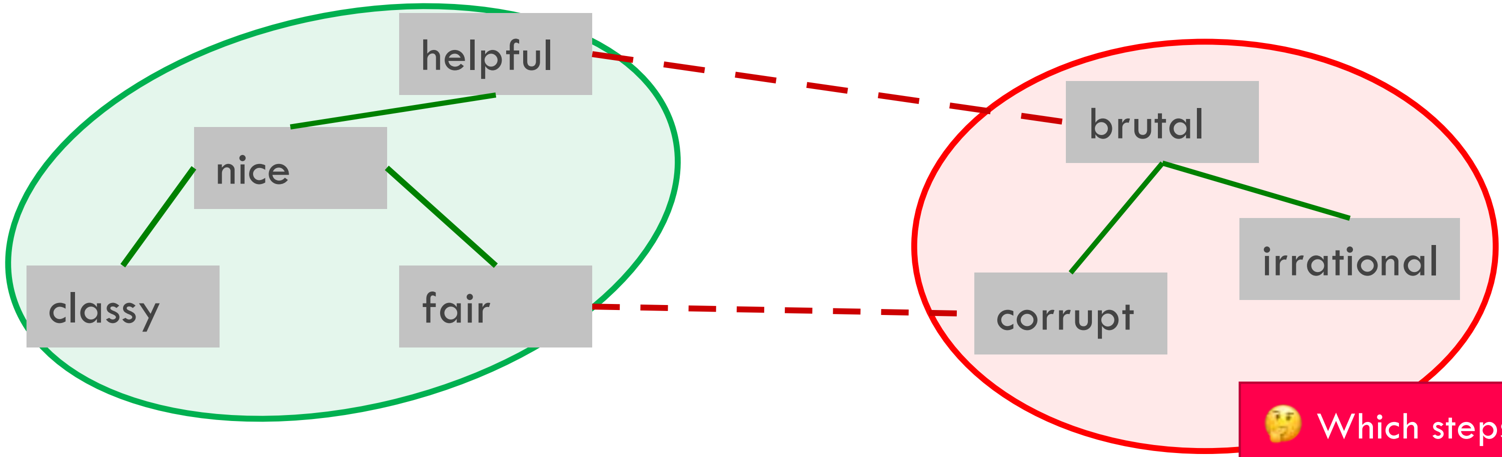
nice, helpful

nice, classy

Slide Credits: Dan Jurafsky (Stanford)

“Snowball” Algorithm

3. Calculate “polarity similarity” for each word pair, resulting in typed edges on a graph



🤔 Which steps are potentially noisy?

4. Cluster the graph into two components.

Slide Credits: Dan Jurafsky (Stanford)

Output polarity lexicon

Positive

- bold decisive disturbing generous good honest important large mature patient peaceful positive proud sound stimulating straightforward strange talented vigorous witty…

Negative

- ambiguous cautious cynical evasive harmful hypocritical inefficient insecure irrational irresponsible minor outspoken pleasant reckless risky selfish tedious unsupported vulnerable wasteful…

Slide Credits: Dan Jurafsky (Stanford)

Using WordNet to learn polarity

We can do the same with an explicit lexicon with links.

1. Create positive and negative seed words (*good*, *terrible*)
2. Find Synonyms and Antonyms

   +ve: Add synonyms of +ve words (*well*) antonyms of –ve words

   –ve: Add synonyms of –ve words, antonyms of +ve words (*evil*)
3. Repeat, following chains of synonyms
4. Filter

Kim and Hovy (2004) *Determining the sentiment of opinions.*
Hu and Liu (2004) *Mining and summarizing customer reviews.*

Summary on Sentiment

Generally modeled as classification or regression task

- Predict a binary or ordinal label

Important features:

- Negation
- All-words incidence (not frequency)
- Leverage lexica of subsets of words
  - Hand-built polarity lexicons
  - Use seeds and semi-supervised learning to induce lexicons

Defining Summarization

Text Summarization

Produce an abridged version of a text that contains information that is important or relevant to a user.

Applications:

- outlines or abstracts of any document, article, etc.
- summaries of email threads
- action items from a meeting
- simplifying text by compressing sentences

Slide Credits: Dan Jurafsky (Stanford)

Dimension 1: What to summarize?

Single-document summarization: Given a document, produce

- Abstract
- Outline
- Headline

Multidocument: Given a group of documents, produce

- a series of news stories on the same event
- a set of web pages about some topic or question

Slide Credits: Dan Jurafsky (Stanford)

2: Query-focused or Generic?

A: Generic summarization:

- Summarize the content of a document

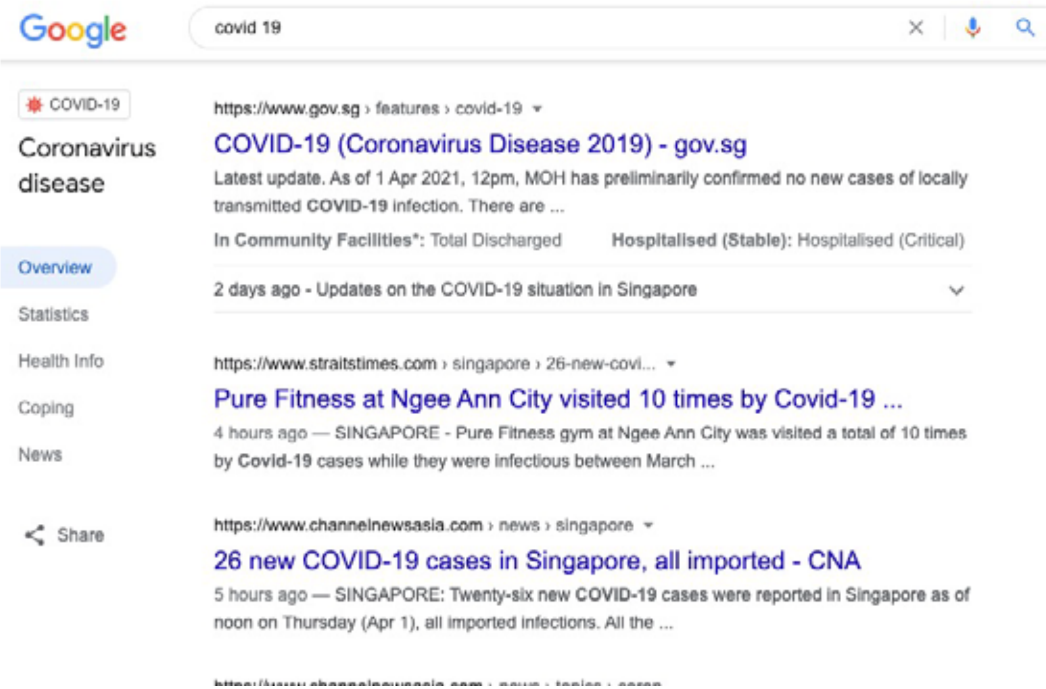
B: Query-focused summarization:

- Summarize a document with respect to an information need expressed in a user query.
- A kind of complex question answering: Answer a question by summarizing a document that has the information to construct the answer.

Slide Credits: Dan Jurafsky (Stanford)

2B1: Summarization for Snippets

Create snippets summarizing a web page for a query



cf. "Passage Retrieval"
Google: 156 characters
(about 26 words) plus title and link

Slide Credits: Dan Jurafsky (Stanford)

2B2: Cohesive Passage

Create answer passage to complex questions summarizing multiple documents.

- Create a single, cohesive answer that combines information from each document, instead of giving a snippet for each document.
- Utilize and remove redundant information.

Slide Credits: Dan Jurafsky (Stanford)

3: Extractive vs. Abstractive

Extractive summarization:

- Create the summary from phrases or sentences in the source document(s)

Edited/Simplified Extraction:

- Fix common problems with compiled extracted sentences

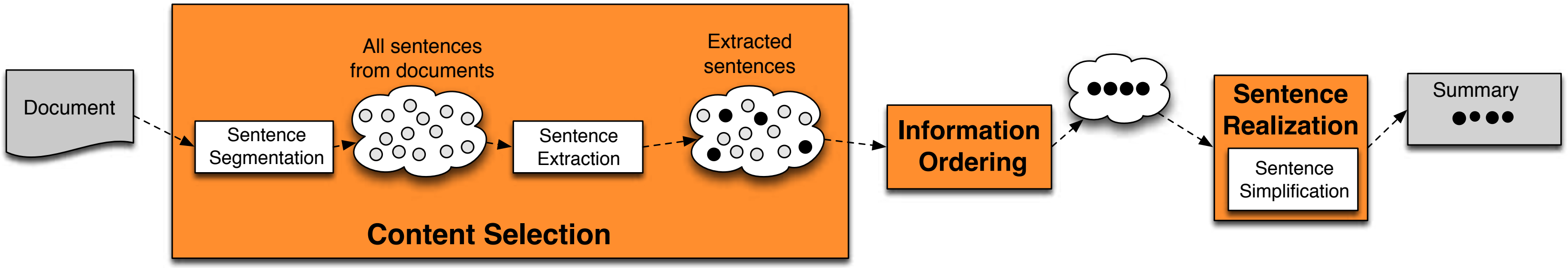
Abstractive summarization:

- Express the ideas in the source documents using (at least in part) different words

Slide Credits: Dan Jurafsky (Stanford)

Summarization Approaches

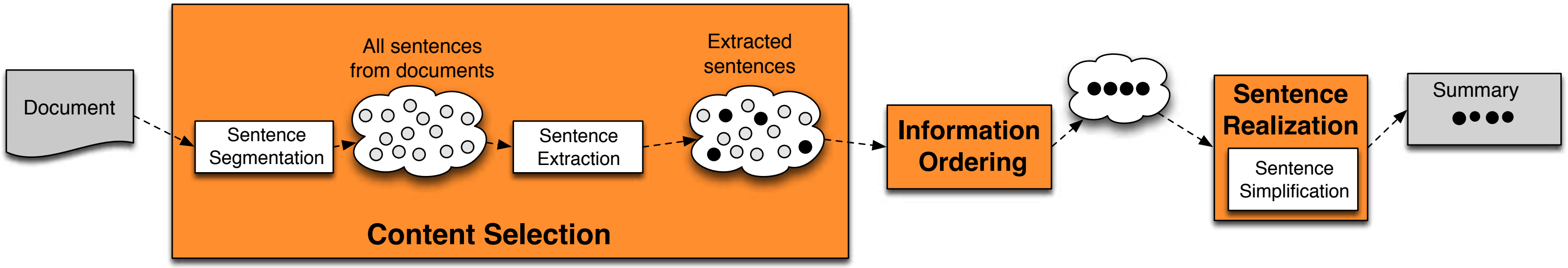
Summarization: Three Stages



1. Content selection: choose sentences to extract
2. Information ordering: choose an order to place them
3. Sentence realization: clean up the sentences

Slide Credits: Dan Jurafsky (Stanford)

Summarization: Basic Algorithm



1. Content selection: choose sentences to extract
2. Information ordering: choose an order to place them
   - Just use the document order (🤔 but what about multidoc?)
3. Sentence realization: clean up the sentences
   - Extractive: keep original sentences

Slide Credits: Dan Jurafsky (Stanford)

Selection baseline: take the first sentence

Google

what is die brucke

All · Images · News · Maps · Videos · More · Settings · Tools

About 2,000,000 results (0.56 seconds)

https://en.wikipedia.org › wiki › Die_Brücke

Die Brücke - Wikipedia

Die Brücke (The **Bridge**) was a group of German expressionist artists formed in Dresden in 1905. ... The seminal group had a major impact on the evolution of modern art in the 20th century and the creation of expressionism. The group came to an end around 1913. The Brücke Museum in Berlin was named after the group.

Ernst Ludwig Kirchner · Erich Heckel · Fritz Bleyl · Karl Schmidt-Rottluff

https://www.theartstory.org › movement › die-brucke

Die Brücke Movement Overview | TheArtStory

21 May 2015 — Summary of **Die Brücke**. Progenitors of the movement later known as German Expressionism, **Die Brücke** formed in Dresden in 1905 as a ...

https://www.britannica.com › Visual Arts › Painting

Die Brücke | art organization | Britannica

Die Brücke, (German: "The Bridge") organization of German painters and printmakers that from 1905 to 1913 played a pivotal role in the development of ...

Die Brücke

From Wikipedia, the free encyclopedia

For other uses, see Die Brücke (disambiguation).

Die Brücke (The Bridge) was a group of German expressionist artists formed in Dresden in 1905. Founding members were Fritz Bleyl, Erich Heckel, Ernst Ludwig Kirchner and Karl Schmidt-Rottluff. Later members were Emil Nolde, Max Pechstein and Otto Mueller. The seminal group had a major impact on the evolution of modern art in the 20th century and the creation of expressionism.[1] The group came to an end around 1913. The Brücke Museum in Berlin was named after the group.

Die Brücke is sometimes compared to the roughly contemporary French group of the Fauves. Both movements shared interests in primitivist art and in the expressing of extreme emotion through high-keyed colors that were very often non-naturalistic. Both movements employed a drawing technique that was crude, and both groups shared an antipathy to complete abstraction. The Die Brücke artists' emotionally agitated paintings of city streets and sexually charged events transpiring in country settings made their French counterparts, the Fauves, seem tame by comparison.[2]

Slide Credits: Dan Jurafsky (Stanford)

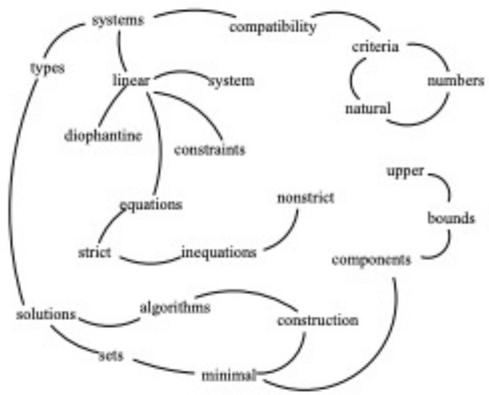
Finding Keywords: Unsupervised Content Selection

Choose sentences that have salient or informative words

How to define salient words?

- Weigh each word w_i in document d_j by its $tf.idf$ (Luhn, 1958)
- Topic signature: choose a smaller set of salient words by its mutual information with the query or its log-likelihood ratio (LLR) (Dunning, 1993; Lin and Hovy, 2000)
- TextRank: top hubs in a word graph, where edges are co-occurences among sentences or other relation types (Mihalcea and Tarau 2004)

Compatibility of systems of linear constraints over the set of natural numbers. Criteria of compatibility of a system of linear Diophantine equations, strict inequations, and nonstrict inequations are considered. Upper bounds for components of a minimal set of solutions and algorithms of construction of minimal generating sets of solutions for all types of systems are given. These criteria and the corresponding algorithms for constructing a minimal supporting set of solutions can be used in solving all the considered types systems and systems of mixed types.



Keywords assigned by TextRank:
linear constraints; linear diophantine equations; natural numbers; nonstrict inequations; strict inequations; upper bounds

Keywords assigned by human annotators:
linear constraints; linear diophantine equations; minimal generating sets; non-strict inequations; set of natural numbers; strict inequations; upper bounds

Slide Credits: Dan Jurafsky (Stanford) Picture Credit: Mihalcea and Tarau 2004)

Evaluating Summaries: ROUGE

ROUGE: n-gram overlap

ROUGE (Recall Oriented Understudy for Gisting Evaluation; Lin and Hovy 2003): Intrinsic metric for automatically evaluating summaries

- Not as good as human evaluation ("Did this answer the user's question?")
- But statistically proven as a much more convenient proxy

Given a document d, and an automatic summary $\hat{y}$:

1. Have n humans produce a set of reference summaries $Y = \{y_1, \dots, y_n\}$
2. What percentage of the bigrams from appear in x?

Slide Credits: Dan Jurafsky (Stanford)

ROUGE: n-gram overlap

ROUGE (Recall Oriented Understudy for Gisting Evaluation; Lin and Hovy 2003): Intrinsic metric for automatically evaluating summaries

- Given a document d, a set of reference summaries $Y = \{y_1, \ldots, y_n\}$ and an automatic summary $\hat{y}$, what percentage of the n-grams from Y appear in $\hat{y}$?

- Not as good as human evaluation ("Did this answer the user's question?")
- But statistically proven as a much more convenient proxy

Slide Credits: Dan Jurafsky (Stanford)

ROUGE Example

Human Answers to the question *What is water spinach?*

y_1: *Water spinach is a green leafy vegetable grown in the tropics.*

y_2 : *Water spinach is a semi-aquatic tropical plant grown as a vegetable.*

y_3: *Water spinach is a commonly eaten leaf vegetable of Asia.*

System answer ($\hat{y}$): *Water spinach is a leaf vegetable commonly eaten in tropical areas of Asia.*

ROUGE-2 =

Slide Credits: Dan Jurafsky (Stanford)

ROUGE Example

Human Answers to the question *What is water spinach?*

y_1: *Water spinach is a green leafy vegetable grown in the tropics.*

y_2 : *Water spinach is a semi-aquatic tropical plant grown as a vegetable.*

y_3: *Water spinach is a commonly eaten leaf vegetable of Asia.*

System answer ($\hat{y}$): *Water spinach is a leaf vegetable commonly eaten in tropical areas of Asia.*

$$\text{ROUGE-2} = \frac{3+3+6}{10+9+9} = 12/28 = 43\%$$

Slide Credits: Dan Jurafsky (Stanford)

Query Based Summarization

Bridging to Question Answering

Answering harder questions: Query-focused multi-document summarization

The (bottom-up) snippet method

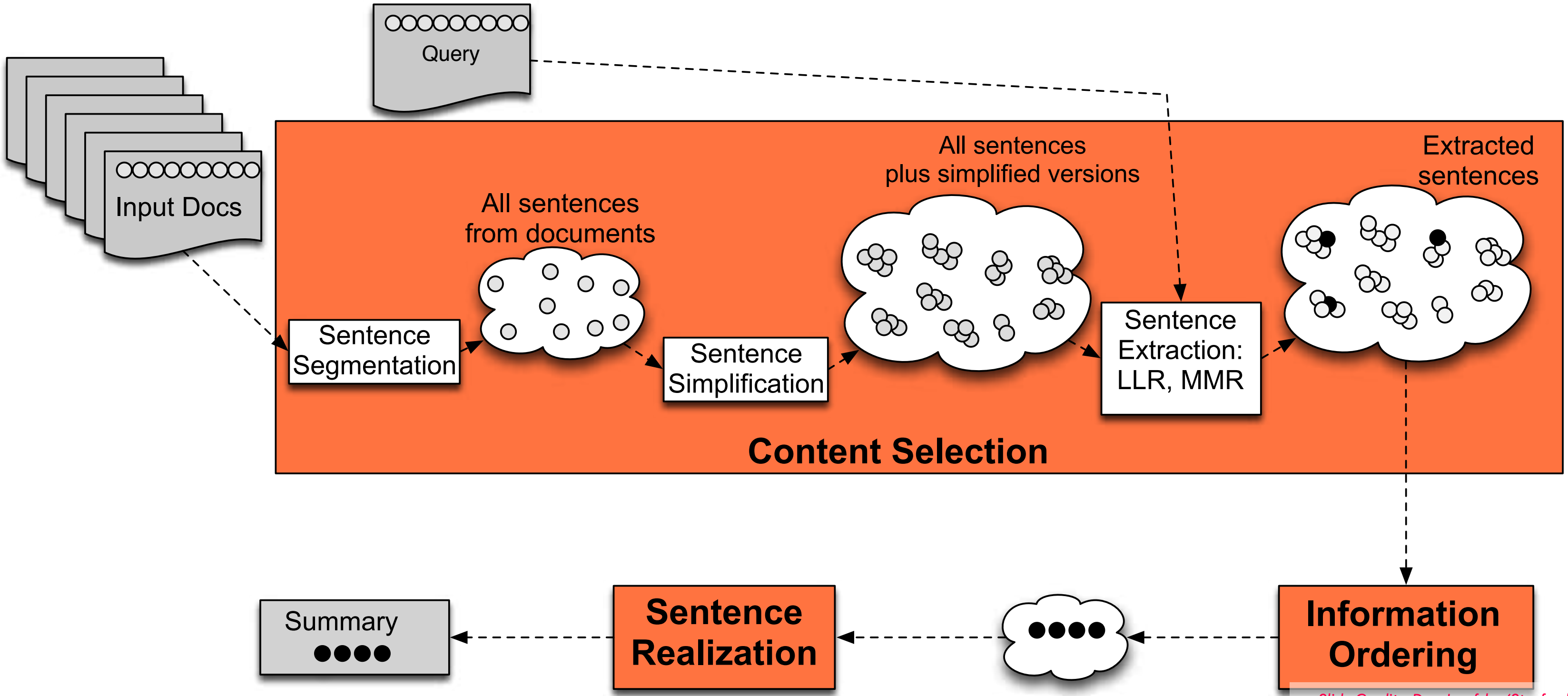
- Find a set of relevant documents
- Extract informative sentences from the documents
- Order and modify the sentences into an answer

The (top-down) information extraction method

- Build specific answerers for different question types:
  - definition questions
  - biography questions
  - certain medical questions

Slide Credits: Dan Jurafsky (Stanford)

Query-Focused Multi-Document Summarization



Slide Credits: Dan Jurafsky (Stanford)

Simplifying sentences

Zajic et al. (2007), Conroy et al. (2006), Vanderwende et al. (2007)

Parse sentences, then use (learned) rules to prune modifiers.

appositives	Rajam, ~~28, an artist who was living at the time in Philadelphia~~, found the inspiration in the back of city magazines.
attribution clauses	Rebels agreed to talks with government officials, ~~international observers said Tuesday.~~
PPs without named entities	The commercial fishing restrictions in Washington will not be lifted unless the salmon population increases [~~PP to a sustainable number~~]]
initial adverbials	"~~For example~~", "~~On the other hand~~", "~~As a matter of fact~~", "~~At this point~~"

Slide Credits: Dan Jurafsky (Stanford)

Maximal Marginal Relevance (MMR)

Content selection from multiple documents

Iteratively, greedily pick the best sentence to add to the existing summary:

- Relevant: Maximally relevant to the user's query
  - high (cosine) similarity to the query
- Novel: Minimally redundant with the existing summary/answer so far
  - low similarity to the summary

$$s_{MMR} = \max_{s \in D} \left[\alpha \cdot sim(s, Q) - (1 - \alpha) \cdot \max_{s \in S}\left(sim(s, S)\right)\right]$$

Stop when desired length reached

Carbonell and Goldstein (1998) *The Use of MMR, Diversity-based Reranking for Reordering Documents and Producing Summaries*

Slide Credits: Dan Jurafsky (Stanford)

Information Ordering

Chronological ordering:

- Order sentences by the date of the document (for summarizing news). (Barzilay, Elhadad, and McKeown, 2002)

Coherence:

- Choose orderings that make neighboring sentences similar (by cosine).
- Choose orderings in which neighboring sentences discuss the same entity (Barzilay and Lapata, 2007)

Topical ordering

- Learn the ordering of topics in the source documents

Slide Credits: Dan Jurafsky (Stanford)

Domain-specific answering: Using Information Extraction

a biography of a person contains:

- person's birth/death, fame factor, education, nationality and so on

a definition contains:

- genus or hypernym: *The Hajj is a type of ritual*

a medical answer about a drug's use contains:

- the problem (the medical condition),
- the intervention (the drug or procedure),
- The comparison (e.g., control group),
- the outcome (the result of the study).

= PICO

Slide Credits: Dan Jurafsky (Stanford)

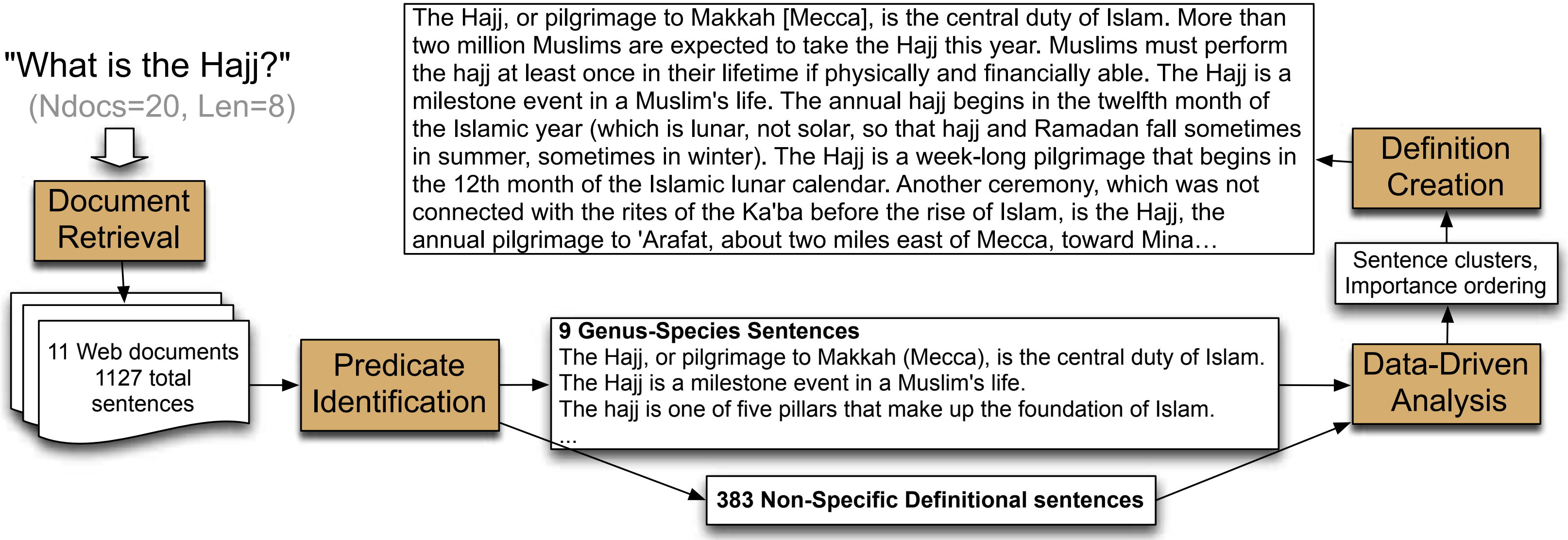
Definitional Templates

	Definition
genus	The Hajj is a type of ritual
species	the annual hajj begins in the twelfth month of the Islamic year
synonym	The Hajj, or Pilgrimage to Mecca, is the central duty of Islam
subtype	Qiran, Tamattu', and Ifrad are three different types of Hajj
	Biography
dates	was assassinated on April 4, 1968
nationality	was born in Atlanta, Georgia
education	entered Boston University as a doctoral student
	Drug efficacy
population	37 otherwise healthy children aged 2 to 12 years
problem	acute, intercurrent, febrile illness
intervention	acetaminophen (10 mg/kg)
outcome	ibuprofen provided greater temperature decrement and longer duration of antipyresis than acetaminophen when the two drugs were administered in approximately equal doses

Slide Credits: Dan Jurafsky (Stanford)

Sample Architecture for complex question answering: definition questions

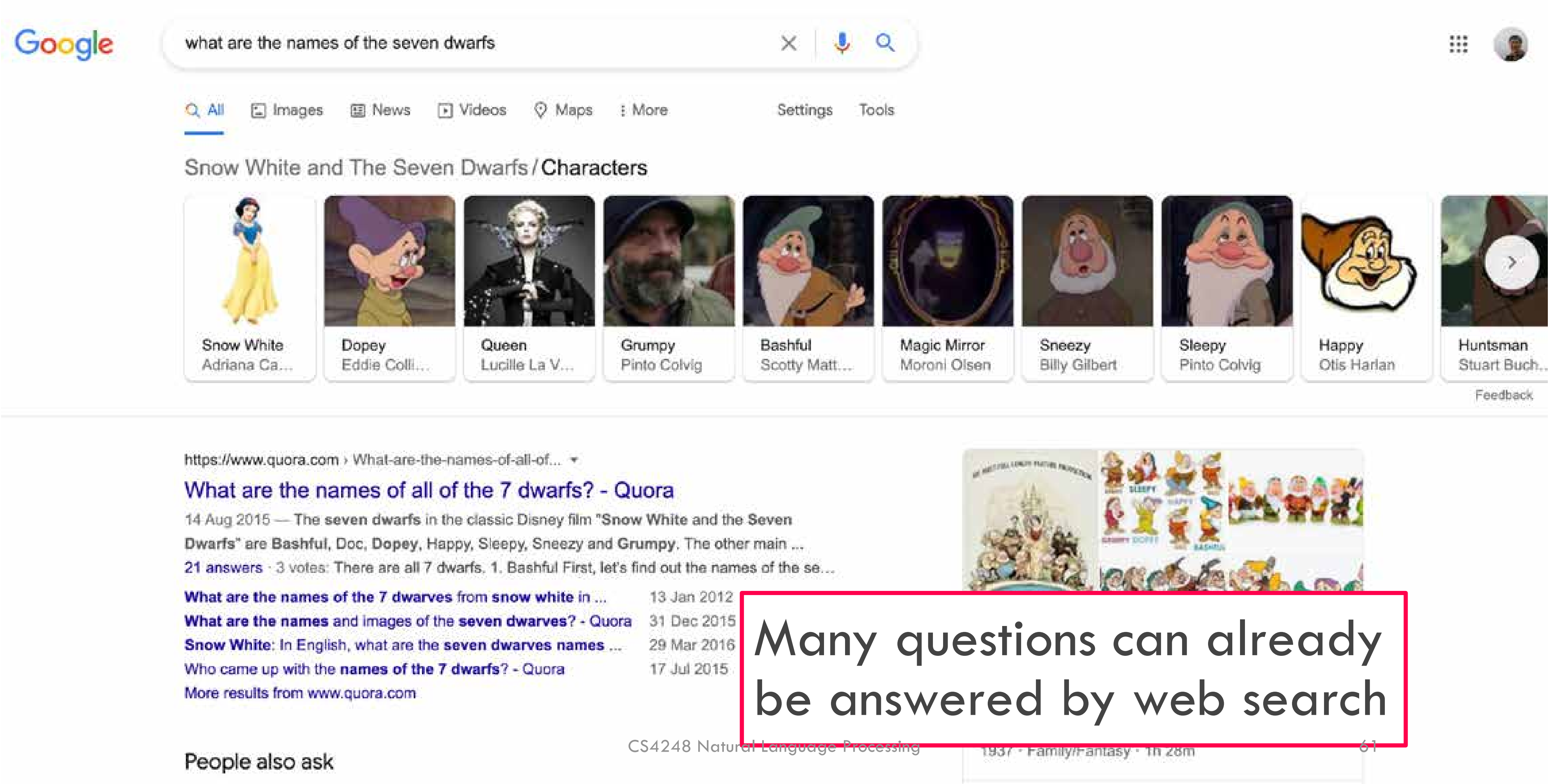
Blair-Goldensohn et al. (2004). *Answering Definition Questions: A Hybrid Approach*.



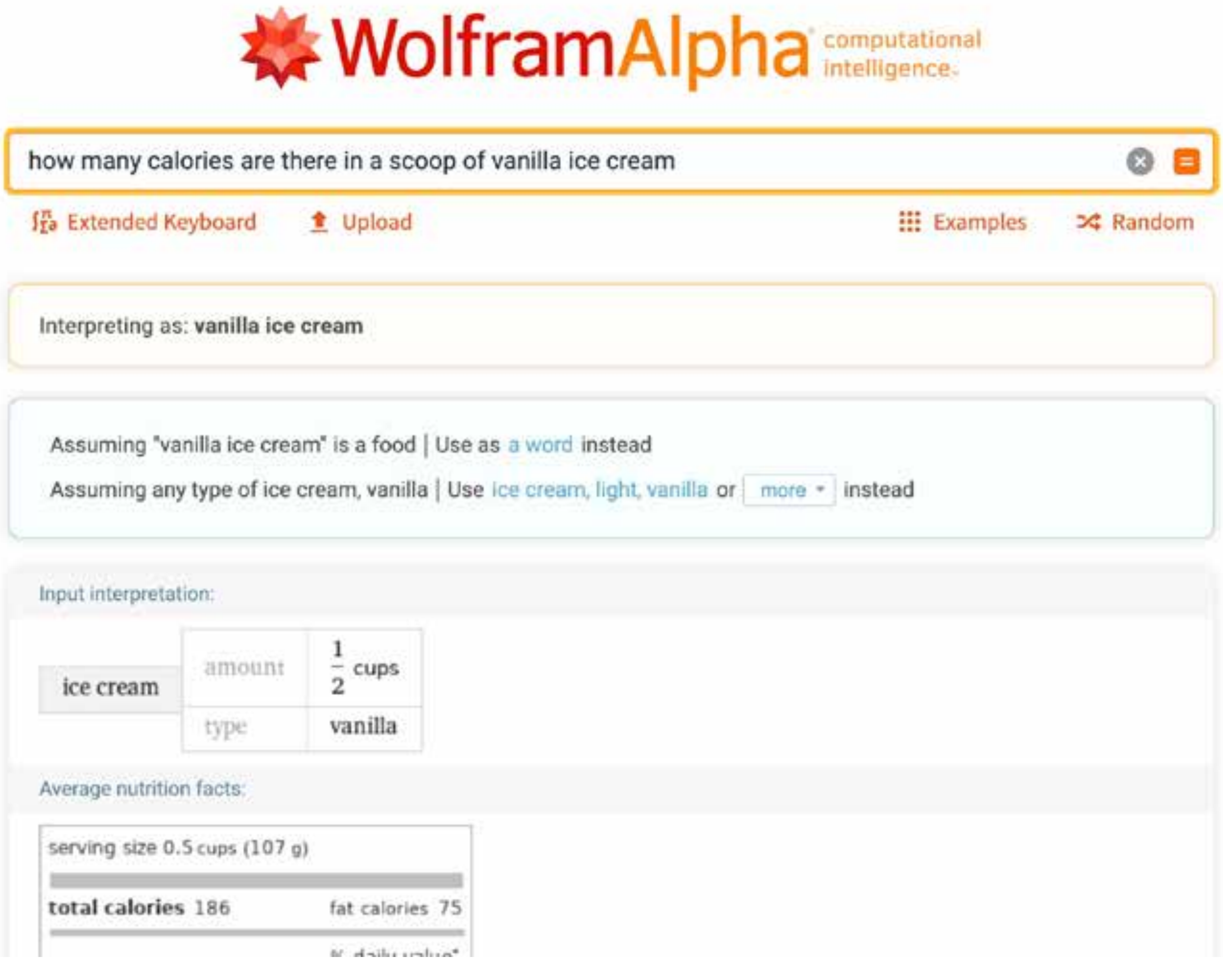
Slide Credits: Dan Jurafsky (Stanford)

Question Answering

IBM Watson won the Jeopardy! trivia competition on February 16, 2011!



Many questions can already be answered by web search


WolframAlpha computational intelligence.
how many calories are there in a scoop of vanilla ice cream
Extended Keyboard
Upload
Examples
Random
Interpreting as: vanilla ice cream
Assuming "vanilla ice cream" is a food | Use as a word instead
Assuming any type of ice cream, vanilla | Use ice cream, light, vanilla or more instead
Input interpretation:
ice cream
amount
1/2 cups
type
vanilla
Average nutrition facts:
serving size 0.5 cups (107 g)
total calories 186
fat calories 75

Types of Questions in Modern Systems

Factoid: *Who wrote "The Universal Declaration of Human Rights"?*

Complex (narrative): *What do historians think about Lee Kuan Yew's position on racial harmony?*

Answers

- A short span of text / A paragraph
- Yes/No
- A database entry
- A list

Context

- A passage, a document, a large collection of documents
- Knowledge base
- Semi-structured tables
- Images / Video

Slide Credits: Diyi Yang (GeorgiaTech)

n)

School of Computing

The first recorded travels by Europeans to China and back date from this time. The most famous traveler of the period was the Venetian Marco Polo, whose account of his trip to "Cambaluc," the capital of the Great Khan, and of life there astounded the people of Europe. The account of his travels, Il milione (or, The Million, known in English as the Travels of Marco Polo), appeared about the year 1299. Some argue over the accuracy of Marco Polo's accounts due to the lack of mentioning the Great Wall of China, tea houses, which would have been a prominent sight since Europeans had yet to adopt a tea culture, as well the practice of foot binding by the women in capital of the Great Khan. Some suggest that Marco Polo acquired much of his knowledge through contact with Persian traders since many of the places he named were in Persian.

How did some suspect that Polo learned about China instead of by actually visiting it?

Answer: through contact with Persian traders

(Rajpurkar et al, 2016): *SQuAD: 100,000+ Questions for Machine Comprehension of Text*

Answer: 2016

- (passage, question, answer) triples
- Passage is from Wikipedia, question is crowd-sourced
- Answer must be a span of text in the passage (aka. "extractive question answering")
- SQuAD 1.1: 100k answerable questions, SQuAD 2.0: another 50k unanswerable questions

Slide Credits: Diyi Yang (GeorgiaTech)

Reading Comprehension MCQ

James the Turtle was always getting in trouble. Sometimes he'd reach into the freezer and empty out all the food. Other times he'd sled on the deck and get a splinter. His aunt Jane tried as hard as she could to keep him out of trouble, but he was sneaky and got into lots of trouble behind her back.

One day, James thought he would go into town and see what kind of trouble he could get into. He went to the grocery store and pulled all the pudding off the shelves and ate two jars. Then he walked to the fast food restaurant and ordered 15 bags of fries. He didn't pay, and instead headed home.

His aunt was waiting for him in his room. She told James that she loved him, but he would have to start acting like a well-behaved turtle.

After about a month, and after getting into lots of trouble, James finally made up his mind to be a better turtle.

1) What is the name of the trouble making turtle?

A) Fries

B) Pudding

C) James

D) Jane

2) What did James pull off of the shelves in the grocery store?

A) pudding

B) fries

C) food

D) splinters

(Richardson et al, 2013): *MCTest: A Challenge Dataset for the Open-Domain Machine Comprehension of Text*

Slide Credits: Diyi Yang (GeorgiaTech)

The Virginia governor's race, billed as the marquee battle of an otherwise anticlimactic 2013 election cycle, is shaping up to be a foregone conclusion. Democrat Terry McAuliffe, the longtime political fixer and moneyman, hasn't trailed in a poll since May. Barring a political miracle, Republican Ken Cuccinelli will be delivering a concession speech on Tuesday evening in Richmond. In recent ...

Q: What are the candidates **running** for? — A: Governor

Q: **Where?** — A: Virginia

Q: Who is the democratic candidate? — A: Terry McAuliffe

Q: Who is **his** opponent? — A: Ken Cuccinelli

Q: What party does **he** belong to? — A: Republican

Q: Which of **them** is winning?

Q: Who is **his** opponent? — A: Ken Cuccinelli

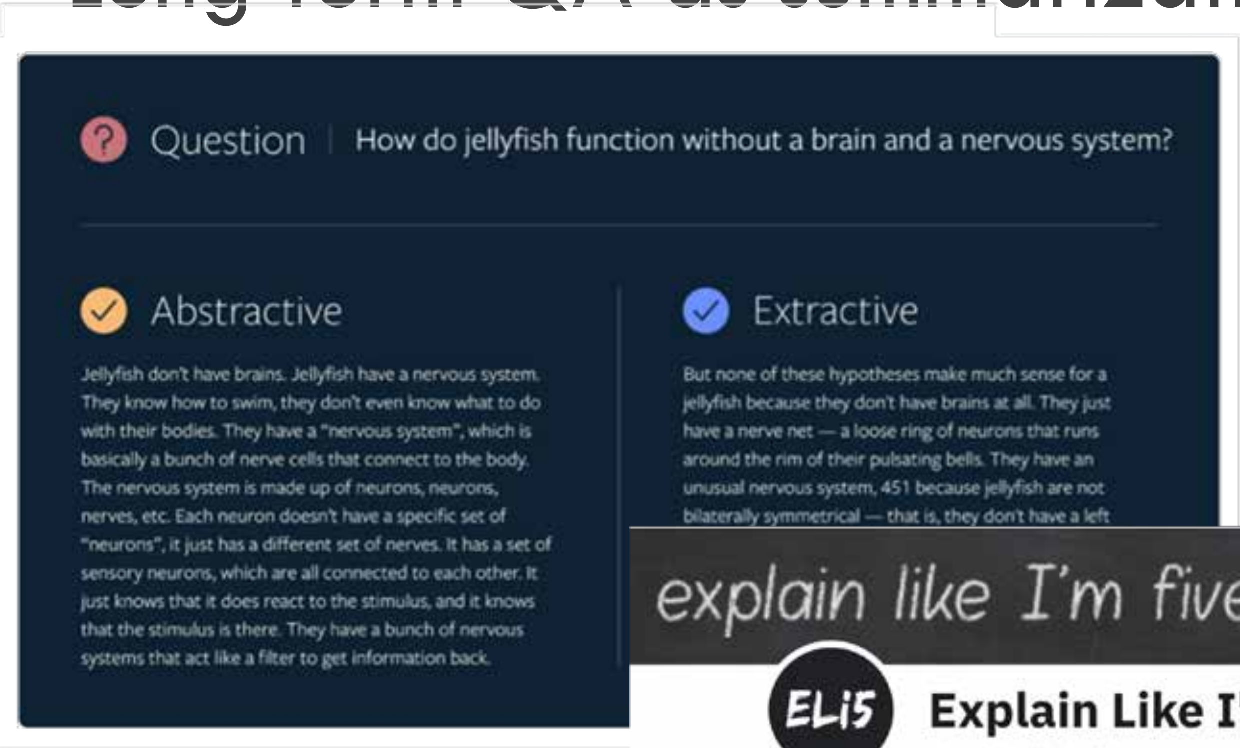
Q: What party does **he** belong to? — A: Republican

Q: Which of **them** is winning?

(Reddy et al, 2019): *CoQA: A Conversational Question Answering Challenge*

Slide Credits: Diyi Yang (GeorgiaTech)

Long form QA as summarization



https://ai.facebook.com/blog/longform-qa/
(Fan et al, 2019): *ELI5: Long Form Question Answering*

Slide Credits: Diyi Yang (GeorgiaTech)

Paradigms for QA

Information Retrieval Based approaches

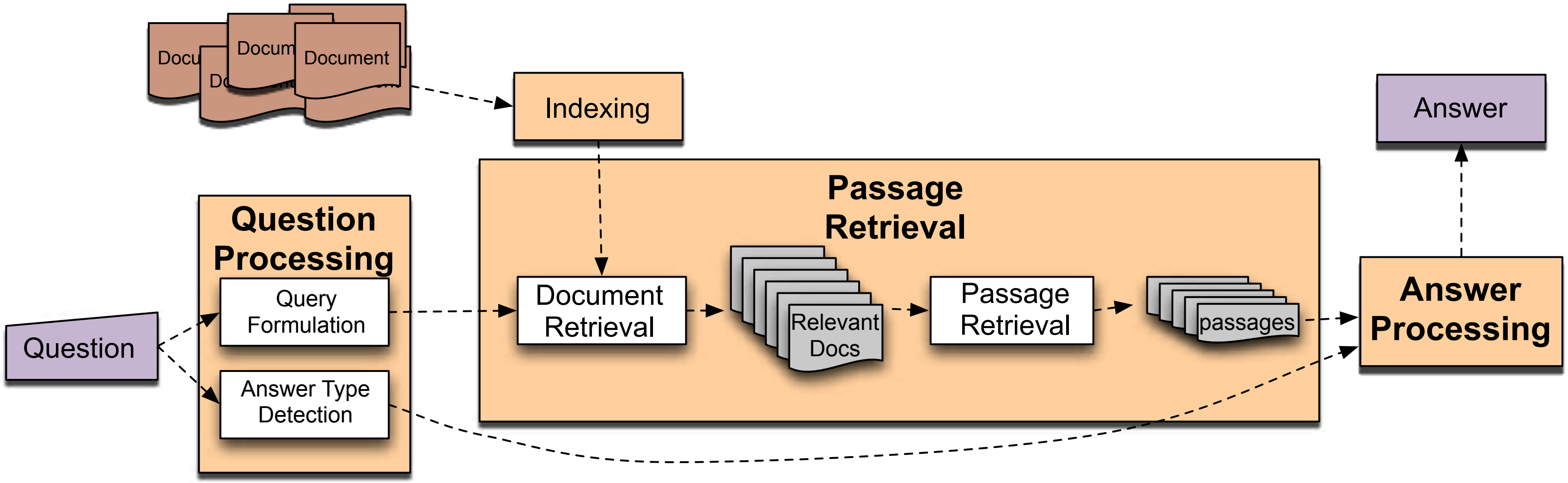
- TREC; IBM Watson; Google

Knowledge Based and Hybrid approaches

- IBM Watson; Apple Siri; Wolfram Alpha; True Knowledge Evi

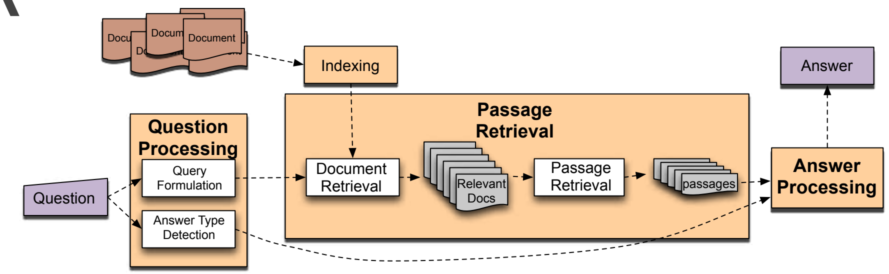
Slide Credits: Dan Jurafsky (Stanford)

IR Based Factoid QA



Slide Credits: Dan Jurafsky (Stanford)

IR Based Factoid QA



1. Question Processing
   - Detect question type, answer type, focus, relations
   - Formulate queries to send to a search engine / database
2. Passage Retrieval
   - Retrieve ranked documents
   - Break into suitable passages and rerank
3. Answer Processing
   - Extract candidate answers
   - Rank candidates using evidence from the text and external sources

Slide Credits: Dan Jurafsky (Stanford)

Knowledge-based approaches (e.g., Siri)

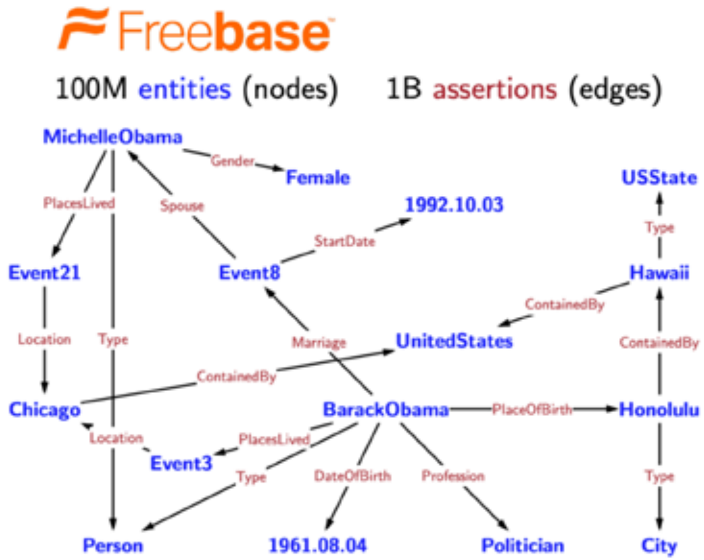
1. Build a semantic representation of the query
    - Times, dates, locations, entities, numeric quantities

2. Map from this semantics to query structured data or resources
    - Geospatial databases
    - Ontologies (Wikipedia Infoboxes, dbPedia, WordNet, Yago)
    - Restaurant review sources and reservation services
    - Scientific databases

Slide Credits: Dan Jurafsky (Stanford)

Freebase QA

(Berant et al, 2013): *Semantic Parsing on Freebase from Question-Answer Pairs*



Slide Credits: Diyi Yang (GeorgiaTech)

Hybrid approaches (IBM Watson)

1. Build a shallow semantic representation of the query
2. Generate answer candidates using IR methods
   - Augmented with ontologies and semi-structured data
3. Score each candidate using richer knowledge sources
   - Geospatial databases
   - Temporal reasoning
   - Taxonomical classification

Slide Credits: Dan Jurafsky (Stanford)

Answer Types and Query Formulation

Question Processing

Things to extract from the question:

Answer Type Detection
- Decide the named entity type (person, place) of the answer

Query Formulation
- Choose query keywords for the IR system

Question Type classification
- Is this a definition question, a math question, a list question?

Focus Detection
- Find the question words that are replaced by the answer

Relation Extraction
- Find relations between entities in the question

Slide Credits: Dan Jurafsky (Stanford)

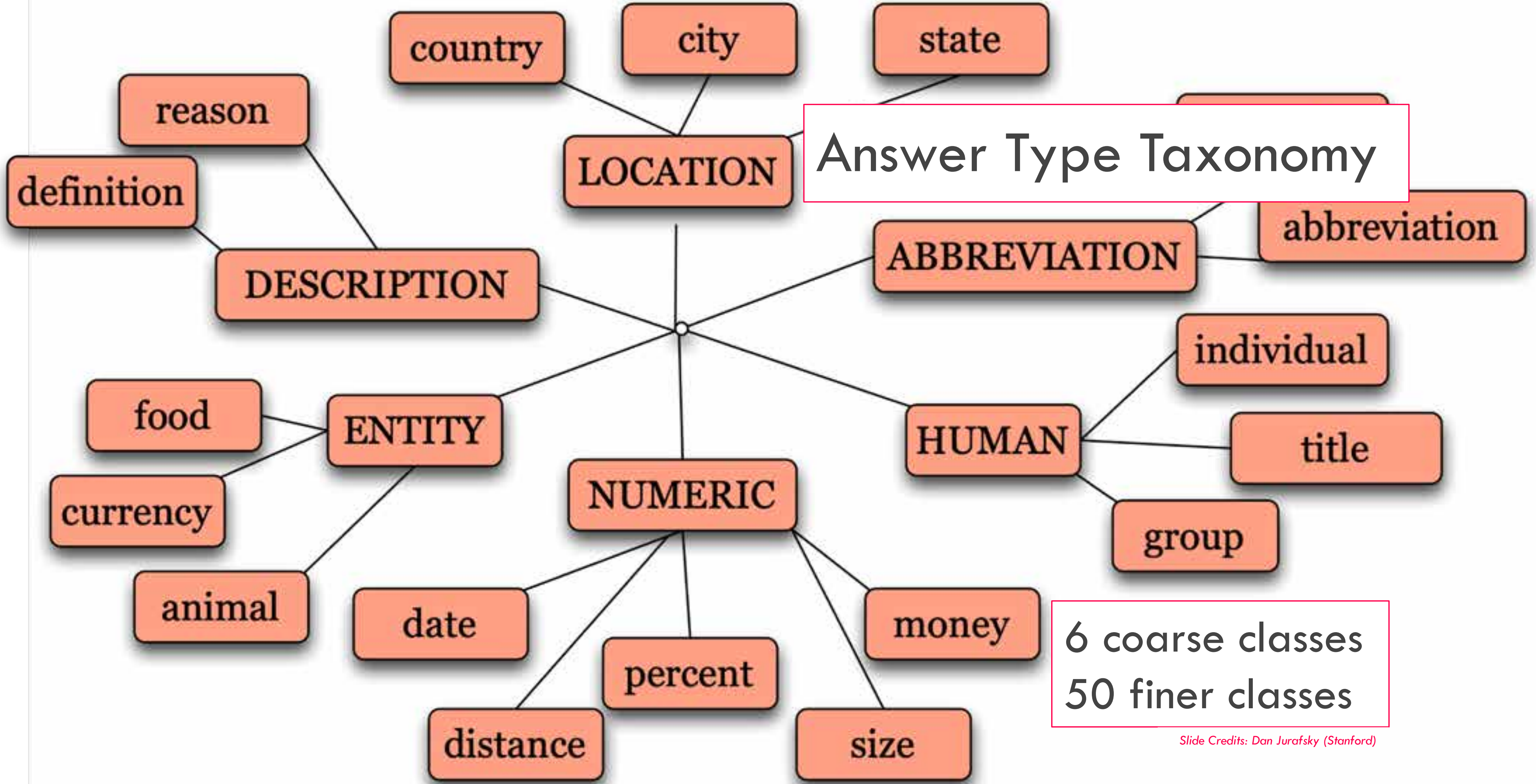
Question Processing

They're the two states you could be re-entering if you're crossing Florida's northern border

- Answer Type: US state
- Query: *two states*, *border*, *Florida*, *north*
- Focus: *the two states*
- Relations: Borders(Florida, λx, north)

Slide Credits: Dan Jurafsky (Stanford)

Answer Type Taxonomy



6 coarse classes
50 finer classes

Slide Credits: Dan Jurafsky (Stanford)

Answer Type Detection

- Regular expression-based rules:
  Who {is|was|are|were} PERSON
  PERSON (YEAR – YEAR)
- Question headword: the headword of the first noun phrase after the *wh*-word)
  Which city in China has the largest number of foreign financial companies?
  What is the state flower of California?

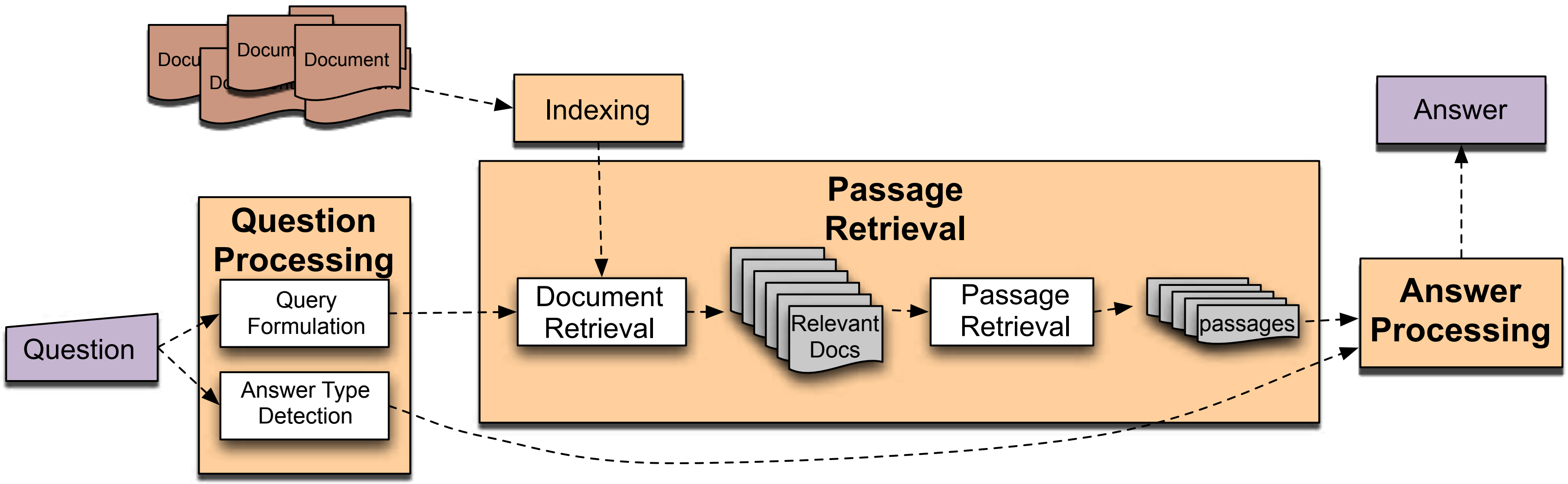
Slide Credits: Dan Jurafsky (Stanford)

Answer type classification as supervised ML

1. Define a taxonomy of question types
2. Annotate training data for each question type
3. Train classifiers using a set of features
   - Question words and phrases
   - Part-of-speech tags
   - Parse features (headwords)
   - Named Entities
   - Semantically related words

Slide Credits: Dan Jurafsky (Stanford)

IR Based Factoid QA



Slide Credits: Dan Jurafsky (Stanford)

Send your Query: Keyword Selection

1. Select all non-stop words in quotations
2. Select all NNP words in recognized named entities
3. Select all complex nominals with their adjectival modifiers
4. Select all other complex nominals
5. Select all nouns with their adjectival modifiers
6. Select all other nouns
7. Select all verbs
8. Select all adverbs
9. Select the QFW word

(skipped in all previous steps)

10. Select all other words

Moldovan et al. (1999) *The Structure and Performance of an Open-Domain Question Answering System*

Slide Credits: Dan Jurafsky (Stanford)

Choosing keywords from the query

~~Who~~ coined ~~the~~ term "cyberspace" ~~in his~~ novel "Neuromancer"?

1 1

4 4

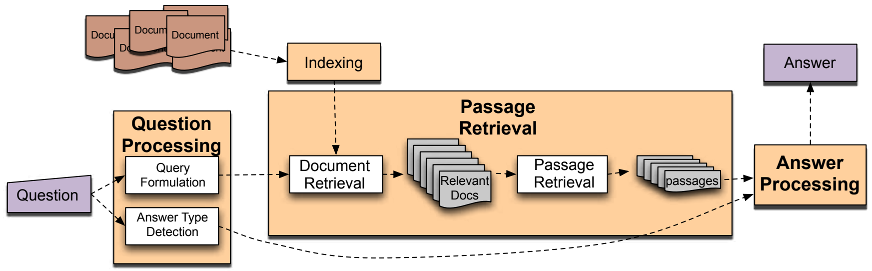
7

cyberspace/1 Neuromancer/1 term/4 novel/4 coined/7

Slide from Mihai Surdeanu

Passage Retrieval and Answer Extraction

Passage Retrieval



1. IR engine retrieves documents using query terms
2. Segment the documents into shorter units
   - Something like paragraphs
3. Passage ranking
   - Use answer type to help re-rank passages

Slide Credits: Dan Jurafsky (Stanford)

NUS National University of Singapore | School of Computing

Features for Passage Ranking

- Number of Named Entities of the right type in passage
- Number of query words in passage
- Number of question N-grams also in passage
- Proximity of query keywords to each other in passage
- Longest sequence of question words
- Rank of the document containing passage

Slide Credits: Dan Jurafsky (Stanford)

Answer Extraction

Run an answer-type named-entity tagger on the passages

- Each answer type requires a named-entity tagger that detects it
- If answer type is CITY, tagger has to tag CITY
  - Can be full NER, simple regular expressions, or hybrid

Return the string with the right type:

- `Who is the prime minister of India` (PERSON)

  `Manmohan Singh``, Prime Minister of India, had told left leaders that the deal would not be renegotiated.`
- `How tall is Mt. Everest?` (LENGTH)

  `The official height of Mount Everest is `**`29035 feet`**

Slide Credits: Dan Jurafsky (Stanford)

NUS National University of Singapore | School of Computing

Features for ranking candidate answers

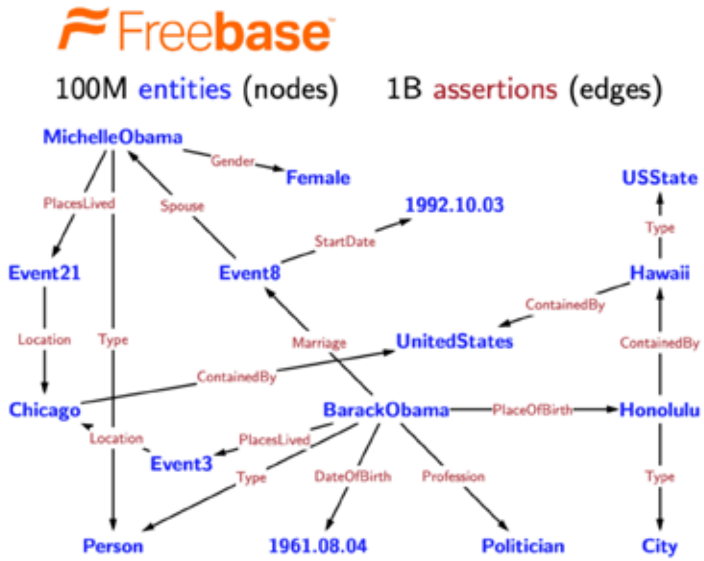
- **Answer type match:** Candidate contains a phrase with the correct answer type.
- **Pattern match:** Regular expression pattern matches the candidate.
- **Question keywords:** # of question keywords in the candidate.
- **Keyword distance:** Distance in words between the candidate and query keywords
- **Novelty factor:** A word in the candidate is not in the query.
- **Apposition features:** The candidate is an appositive to question terms
- **Punctuation location:** The candidate is immediately followed by a comma, period, quotation marks, semicolon, or exclamation mark.
- **Sequences of question terms:** The length of the longest sequence of question terms that occurs in the candidate answer.

Slide Credits: Dan Jurafsky (Stanford)

Using Knowledge in QA

Recap: Freebase QA

Berant et al. (2013): *Semantic Parsing on Freebase from Question-Answer Pairs*



Slide Credits: Diyi Yang (GeorgiaTech)

Relation Extraction

Answers: Databases of Relations

- *bornIn*("Emma Goldman", "June 27 1869")
- *authorOf*("Cao Xue Qin", "Dream of the Red Chamber")
- Draw from Wikipedia infoboxes, DBpedia, FreeBase, etc.

Questions: Extracting Relations in Questions

Whose granddaughter starred in E.T.?

$\Longrightarrow$ (actedIn λx "E.T.")

(granddaughterOf λx λy)

Slide Credits: Dan Jurafsky (Stanford)

Temporal Reasoning

Use Relational Databases

(and obituaries, biographical dictionaries, etc.)

Example from IBM Watson:

"*In 1594 he took a job as a tax collector in Andalusia*"

- Candidates:
  - Thoreau is a bad answer (born in 1817)
  - Cervantes is possible (was alive in 1594)

Slide Credits: Dan Jurafsky (Stanford)

Context and Conversation in Virtual Assistants

Coreference helps resolve ambiguities

👩 : "*Book a table at Il Fornaio at 7:00 with my mom*"

👩 : "*Also send her an email reminder*"

Clarification questions:

👩: "*Chicago pizza*"

🤖: "*Did you mean pizza restaurants in Chicago or Chicago-style pizza?*"

Slide Credits: Dan Jurafsky (Stanford)

QA Evaluation

Common Evaluation Metrics

Accuracy (does answer match gold-labeled answer?)

Mean Reciprocal Rank (MRR)

- For each query, return a ranked list of M candidate answers.
- Query score is $1/Rank$ of the first correct answer
  - If first answer is correct: 1
  - else if second answer is correct: ½
  - else if third answer is correct: ⅓, etc.
  - Score is 0 if none of the M answers are correct
- Take the mean over all n queries

Slide Credits: Dan Jurafsky (Stanford)

Summary: Classification Applications

Many classification tasks becomes supervised machine learning

They can be accompanied by the definition of good feature classes (rather than individual features)

Weight these features using a classifier

Manipulate natural language to engineer features and lexicons for use in tasks

Click to edit Master Attribution style.