

Introduction to Natural Language Processing

Vered Shwartz

Bar-Ilan University, Israel

June 13, 2018

TMI: Too Much Information

Let's start with two facts:

TMI: Too Much Information

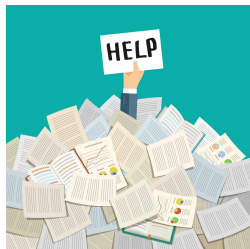
Let's start with two facts:



- ▶ 90% of the data in the world today has been created in the last two years.^[1]

TMI: Too Much Information

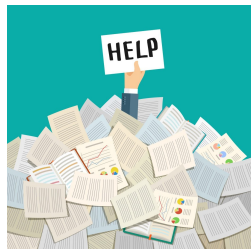
Let's start with two facts:



- ▶ 90% of the data in the world today has been created in the last two years.^[1]
- ▶ Our attention span is now less than that of a goldfish,^[2] and we almost never read through an article.^[3]

TMI: Too Much Information

Let's start with two facts:



- ▶ 90% of the data in the world today has been created in the last two years.^[1]
- ▶ Our attention span is now less than that of a goldfish,^[2] and we almost never read through an article.^[3]

[1] IBM, March 2012 (!)

[2] The Telegraph, March 2016

[3] Slate, March 2016

TMI: Too Much Information

Why is this a problem?

Sort by: **Relevance** Price

The Greenwich Hotel **\$775**
4.4 ★★★★★ · 5-star hotel
Stylish country-house accommodations that include a lantern-lit indoor pool & dining hot spot.
25.9 mi
Free Wi-Fi

AKA Central Park **\$525**
4.3 ★★★★★ · 4-star hotel
Extended-stay hotel offering studios & suites with full kitchens, plus espresso & cocktail bars.
22.8 mi
Free Wi-Fi

New York Marriott Marquis **\$279**
4.2 ★★★★★ · 4-star hotel
Relaxed rooms in a high-rise, modern hotel offering a revolving restaurant & fitness facilities.
23.2 mi

Kimpton Ink48 Hotel **\$239**
4.2 ★★★★★ · 4-star hotel
Dramatic Hudson River & Midtown skyscraper views from spacious rooms & rooftop Press Lounge.
22.8 mi
Free Wi-Fi

Trump International Hotel & Tower New York **\$447**
4.4 ★★★★★ · 5-star hotel
High-end rooms with floor-to-ceiling windows, plus 2 restaurants, a terrace & an indoor pool.
22.5 mi

Washington Square Hotel **\$255**
4.2 ★★★★★ · 3-star hotel
Refined parkside property with art deco-style guest rooms, plus a restaurant, lobby bar & free WiFi.
25.0 mi
Free Wi-Fi Free breakfast

Central Park **\$114**

Service & facilities · 4.0 ★★★★★
Guests appreciated the friendly staff - Some guests said management could be improved

Google reviews

Mitch Dick
4 weeks ago
★☆☆☆☆ Terrible Experience!!!! Our bathroom was gross, hair all over the shower and the toilet clearly wasn't cleaned. Was over charged \$700.00 from a Bar tab that didn't belong to me. Charged me twice for Parking. Took over 3 weeks to get ... [More](#)

Jeff Curry
3 weeks ago
★★★★★ Wonderful hotel with clean rooms and friendly staff. We stay here every year for the Five Boro Bike Tour and they are very accommodating with our bikes. Any problems are presented with transparency and dealt with in a timely manner. They ... [More](#)

Wayne Williams
a week ago
★★★★☆ The rooms are standard for a hotel of this sort the WiFi however has been incredibly slow. if you are platinum you plus 1 guest have access to concierge lounge and they are strict on the 1 guest. The TV does not have HDTV which I thought ... [More](#)

[More Google reviews](#)

Reviews from the web

B Booking.com **3,210** · 5,419 reviews **P** priceline.com **9,019** · 902 reviews

- 7.0** All, 4 weeks ago
★ The location and services.
✖ Some of the staff at front desk are not friendly
- 3.0** Fahad, a month ago
★ Nothing.
✖ Staff was rude. hotel was Not clean. they deducted the amount first and later told me to wait since no rooms available. made me wait for three hours ... [Full review](#)
- 8.3** Tina, a month ago
★ Brilliant location right in the middle of all the hustle and bustle of Time Square. Huge hotel we stayed on the 27th floor. Nice family room with ...
✖ Only free wifi in the lobby really expensive to add on. Should be included as it's not the cheapest place to stay. Bar very expensive \$19 for a ... [Full review](#)

[View 5,419 verified Booking.com reviews](#)

TMI: Too Much Information

Why is this a problem?

Sort by: **Relevance** Price

The Greenwich Hotel **\$775**
 4.4 ★★★★★ · 5-star hotel
 Stylish country-house accommodations that include a lantern-lit indoor pool & dining hot spot.
 25.9 mi
 Free Wi-Fi

AKA Central Park **\$525**
 4.3 ★★★★★ · 4-star hotel
 Extended-stay hotel offering studios & suites with full kitchens, plus espresso & cocktail bars.
 22.8 mi
 Free Wi-Fi

New York Marriott Marquis **\$279**
 4.2 ★★★★★ · 4-star hotel
 Relaxed rooms in a high-rise, modern hotel offering a revolving restaurant & fitness facilities.
 23.2 mi

Kimpton Ink48 Hotel **\$239**
 4.2 ★★★★★ · 4-star hotel
 Dramatic Hudson River & Midtown skyscraper views from spacious rooms & rooftop Press Lounge.
 22.8 mi
 Free Wi-Fi

Trump International Hotel & Tower New York **\$447**
 4.4 ★★★★★ · 5-star hotel
 High-end rooms with floor-to-ceiling windows, plus 2 restaurants, a terrace & an indoor pool.
 22.5 mi

Washington Square Hotel **\$255**
 4.2 ★★★★★ · 3-star hotel
 Refined parkside property with art deco-style guest rooms, plus a restaurant, lobby bar & free WiFi.
 25.0 mi
 Free Wi-Fi Free breakfast

Central Park **\$194**

Service & facilities · 4.0 ★★★★★
 Guests appreciated the friendly staff - Guests enjoyed the fitness center - Some guests said management could be improved

Google reviews

M Mitch Dick
 4 weeks ago

★☆☆☆ Terrible Experience!!!! Our bathroom was gross, hair all over the shower and the toilet clearly wasn't cleaned. Was over charged

Yelp Find local, cheap dining, more! Near San Francisco Sign Up Log In

Home About Us Write a Review Find Friends Messages Talk Events

Hilton Garden Inn Tampa/Riverview/Brandon

5 reviews · 1.0 stars

Hotels / San

4329 Garden Vista Dr Riverview, FL 33578
 Get Directions
 (813) 629-6510
 Message the Business
 tampahthead.hgi.com

4329 Garden Vista Dr
 Riverview, FL 33578
 Get Directions
 (813) 629-6510
 Message the Business
 tampahthead.hgi.com

Hotel Exterior

Book a Room
 Show Prices

Today Open 24 hours Open now
 Price range: Moderate

overstock.com FREE SHIPPING OVER \$50

Recommended Reviews

Uptown Brew and BBQ 21.4 miles away from Hilton Garden Inn Tampa/Riverview/Brandon
 Carey F. said "I was extremely impressed with the food, service and the consideration that went into the menu and decor. We had pork..." read more

Elite Events Catering 8.9 miles away from Hilton Garden Inn Tampa/Riverview/Brandon
 Established in 2000, Elite Events Catering has been offering top-production, high-end event planning and catering throughout the entire Tampa... read more

Review

7.0 All The Sor
 3.0 Fish Not Star told reel
 8.3 Time Brill Hug
 6/2/2014
 Great hotel and awesome staff. Ate at Restaurant bar and not the cheapest place to stay. Bar very expensive \$19 for a ... Full review

View 5,419 verified Booking.com reviews

TMI: Too Much Information

Why is this a problem?

Sort by: Relevance Price

The Greenwich Hotel **\$775**
4.4 ★★★★★ 5-star hotel
Stylish country-house accommodations that include a lantern-lit indoor pool & dining hot spot.
25.9 mi
Free Wi-Fi

AKA Central Park **\$525**
4.3 ★★★★★ 4-star hotel
Extended-stay hotel offering studios & suites with full kitchens, plus espresso & cocktail bars.
22.8 mi
Free Wi-Fi

New York Marriott Marquis **\$279**
4.2 ★★★★★ 4-star hotel
Relaxed rooms in a high-rise, modern hotel offering a revolving restaurant & fitness facilities.
23.2 mi

Kimpton Ink48 Hotel **\$239**
4.2 ★★★★★ 4-star hotel
Dramatic Hudson River & Midtown views

Service & facilities 4.0 ★★★★★
Guests appreciated the friendly staff - Some guests enjoyed the fitness center - Some guests said management could be improved

Google reviews

Mitch Dick
4 weeks ago
★★★★★ Terrible Experience!!!! Our bathroom was gross, hair all over the shower and the toilet clearly wasn't cleaned. Was over charged

yelp Find local businesses, view reviews, photos & hours for over 80 million businesses near you. **San Francisco** Sign Up Log In

Hilton Garden Inn Tampa/Riverview/Brandon
4.3 ★★★★★ 5 reviews
\$\$\$ Hotels / San

4328 Garden Vista Dr
Riverview, FL 33578

Book a Room
Show Prices

Today Open 24 hours
Open now
Price range: Moderate

overstock.com FREE SHIPPING OVER \$50

Booking.com online hotel reservations
home

Search Hotels
Where?
Vientiane, Laos

Check-in date: This 1 | December 11 | Check-out date: Sat 3 | December 11 |
I don't have specific dates yet

Rooms 1 | Adults 2 | Children 0

Search

173,348 hotels worldwide

Booking.com on your mobile
Free download

Save 50% or more with Flash Deals
Subscribe for exclusive offers

My viewed hotels

Mercurie Vientiane (Formerly Novotel) ★★★★★
Unit 19 Samsenhar Plaza #12 Box 353, Vientiane, Laos
Score from 14 reviews: above average, 4.9
Latest booking: November 14 at

Win Hotel (Formerly known as Van Huong Hotel) ★★★★★
Chao Anou Road, Ban Anou, Vientiane, Laos
Score from 12 reviews: Pleasant, 6.3
Latest booking: yesterday at

Dave
Mature couple
United Kingdom

Firstly it should be pointed out that you can get a decent room in Chiang Mai for around 500 bahts. But for the equivalent of a night at a Travelodge you can get an excellent room at this

9.2

TMI: Too Much Information

Why is this a problem?

Sort by: Relevance Price

- The Greenwich Hotel** \$775
4.4 ★★★★★ · 5-star hotel
Stylish country-house accommodations that include a lantern-lit indoor pool & dining hot spot.
25.9 mi
Free Wi-Fi
- AKA Central Park** \$525
4.3 ★★★★★ · 4-star hotel
Extended-stay hotel offering studios & suites with full kitchens, plus espresso & cocktail bars.
22.8 mi
Free Wi-Fi
- New York Marriott Marquis** \$279
4.2 ★★★★★ · 4-star hotel
Relaxed rooms in a high-rise, modern hotel offering a revolving restaurant & fitness facilities.
23.2 mi
- Kimpton Ink48 Hotel** \$239
4.2 ★★★★★ · 4-star hotel
Dramatic Hudson River & Manhattan views

Service & facilities · 4.0 ★★★★★
Guests appreciated the friendly staff - Some guests enjoyed the fitness center - Some guests said management could be improved

Google reviews

M Mitch Dick
4 weeks ago
★★★★★ Terrible Experience!!!! Our bathroom was gross, hair all over the shower and the toilet clearly wasn't cleaned. Was over charged

Yelp Hilton Garden Inn Tampa/Riverview/Brandon
4.3 ★★★★★ · 5 reviews
\$5 Hotels

4329 Garden Vista Dr
Riverview, FL 33578

Book a Room
Show Prices

Today Open 24 hours Open now

BOOKING.COM
online hotel reservations

home

Search Hotels
Where?

Check-in date: - | Check-out date: -

I don't have specific dates yet

Rooms: Adults: Children:

173,348 hotels worldwide

Booking.com on your mobile
Free download

Say Hi to Google maps
tripadvisor.com hotel

My view
[Get Directions](#) [My Maps](#) [Edit this place](#) ✓ [Owner-verified listing](#)

Hotel Casa Camper ☆
Carrer d'Elisabets, 11, 08001 Barcelona, Spain
+34 933 42 62 80
[casacamper.com](#)
[Directions](#) [Search nearby](#) [Save to...](#) [more](#)

Categories: Hotel, Boutique Hotel
Book now: 12/19/10 to 12/20/10 \$237 ▼
Sunday Monday

Dave
Mature couple
Talents, United Kingdom

Firstly it should be pointed out that you can get a bahts. But for the equivalent of a night at a Traveleo

TMI: Too Much Information

Why is this a problem?

Sort by: Relevance Price

The Greenwich Hotel 4.4 ★★★★★ · 5-star hotel
Stylish country-house accommodations that include a lantern-lit indoor pool & dining hot spot.
25.9 mi
Free Wi-Fi

AKA Central Park 4.3 ★★★★★ · 4-star hotel
Extended-stay hotel offering studios & suites with full kitchens, plus espresso & cocktail bars.
22.8 mi
Free Wi-Fi

New York Marriott Marquis 4.2 ★★★★★ · 4-star hotel
Relaxed rooms in a high-rise, modern hotel offering a revolving restaurant & fitness facilities.
23.2 mi

Kimpton Ink48 Hotel 4.2 ★★★★★ · 4-star hotel
Dramatic Hudson River & Midtown skyscraper

Service & facilities · 4.0 ★★★★★
Guests appreciated the friendly staff - Guests enjoyed the fitness center
guests said management could be improved

Google reviews

Mitch Dick
4 weeks ago

★☆☆☆ Terrible Experience!!!! Our bathroom was gross the shower and the toilet clearly wasn't cleaned. Was over c

yelp Find local, cheap, dinner, etc.
Home About Us Write a Review

Hilton Garden Inn Tampa/Riverview/Brandon
5 reviews · 4.0 stars
\$5 Hotels · 1.5 mi

4329 Garden Vista Dr
Riverview, FL 33578

"Great Stay"

○○○○○ Review of **The Muse Hotel New York**

CharlesandKelly
Hudson Valley

Our recent stay at the Muse in New York City was superb. We enjoyed the complimentary wine hour and our time at the in house bar prior to our show reservations at Radio City. The room was the basic King room, but was most comfortable, super clean, nicely decorated, and covered all needs. We took advantage of the parking garage, which made our arrival and departure very easy. Upon checking out the staff saw that I had not used my Karma club "raid the bar" credit, so they offered to credit it towards my bar charge. Recommend joining the Karma Club prior to booking to take advantage of free internet access as well. All staff members were professional and pleasant. We will most definitely stay here again. Thanks to all!

BOOKING.COM
online hotel reservations

home

Search Hotels
Where?

Check-in date: - to -

don't have specific dates yet

Rooms: Adults: Children:

173,348 hotels worldwide

Booking.com on your mobile
Free download

Say **Google maps**

tripadvisor.com hotel

My view
[Get Directions](#) [My Maps](#)

Hotel Casa Camper ☆
Carrer d'Elisabets, 11, 08001 Barcelona, Spain
+34 933 42 62 80
[casacamper.com](#)
[Directions](#) [Search nearby](#) [Save to...](#) [more](#)

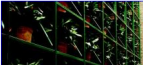
Categories: Hotel, Boutique Hotel
Book now: 12/19/10 to 12/20/10 **\$237** ▼
Sunday Monday

Helpful?

Book a Room

Today Open 24 hours [Open now](#)

[Edit this place](#) ✓ [Owner-verified listing](#)



Dave
Mature couple
Talents, United Kingdom

Firstly it should be pointed out that you can get a bahts. But for the equivalent of a night at a Travele

TMI: Too Much Information

Why is this a problem?

Sort by: Relevance Price

The Greenwich Hotel **\$775**
4.4 ★★★★★ · 5-star hotel
Stylish country-house accommodations that include a lantern-lit indoor pool & dining hot spot.
25.9 mi
Free Wi-Fi

AKA Central Park **\$525**
4.3 ★★★★★ · 4-star hotel
Extended-stay hotel offering studios & suites with full kitchens, plus espresso & cocktail bars.
22.8 mi
Free Wi-Fi

New York Marriott Marquis **\$279**
4.2 ★★★★★ · 4-star hotel
Relaxed rooms in a high-rise offering a revolving restaurant.
23.2 mi

Kimpton Ink48 Hotel
4.2 ★★★★★ · 4-star hotel
Dramatic Hudson River views

Service & facilities · 4.0 ★★★★★
Guests appreciated the friendly staff - Guests enjoyed the fitness center
guests said management could be improved

Google reviews



Mitch Dick
4 weeks ago

★☆☆☆ Terrible Experience!!!! Our bathroom was gross the shower and the toilet clearly wasn't cleaned. Was over c



Hilton Garden Inn Tampa/Riverview/Brandon
5 reviews · 4.0 stars
\$\$ Hotels

"Great Stay"

★★★★★ Review of [The Muse Hotel New York](#)



Charles and Kelly
Hudson Valley

Our recent stay at the Muse in New York City was superb. We enjoyed the complimentary wine hour and our time at the in house bar prior to our show reservations at Radio City. The room was the basic King room, but was most comfortable, super clean, nicely decorated, and covered all needs. We took advantage of the parking garage, which made our arrival and departure very easy. Upon checking out the staff saw that I had not used my Karma club "raid the bar" credit, so they offered to credit it towards my

Some people may spend their entire vacation...
trying to find the optimal hotel!

Search Hotels

Where?

Vientiane, Laos

Check-in date

Thu 1 · December '11

Check-out date

Sat 3 · December '11

I don't have specific dates yet

Rooms 1

Adults 2

Children 0

Search



173,348 hotels worldwide

Dave

Mature couple
Tellers, United States

Firstly it should be pointed out that you can get a bahts. But for the equivalent of a night at a Travele



Booking.com on your mobile

Free download



Say



Google maps

tripadvisor.com hotel

My view

[Get Directions](#) [My Maps](#)

[Edit this place](#) · [Owner-verified listing](#)



Hotel Casa Camper ☆

Carrer d'Elisabets, 11, 08001 Barcelona, Spain

+34 933 42 62 80

[casacamper.com](#)

[Directions](#) [Search nearby](#) [Save to...](#) [more](#)

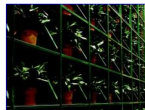
Categories: Hotel, Boutique Hotel

Book now: 12/19/10 to 12/20/10

Sunday

Monday

\$237 ▼



Natural Language Processing to the rescue

We are working on automatic methods to...



- ▶ Summarize multiple long texts

Natural Language Processing to the rescue

We are working on automatic methods to...



- ▶ Summarize multiple long texts
- ▶ Answer questions based on texts

Natural Language Processing to the rescue

We are working on automatic methods to...



- ▶ Summarize multiple long texts
- ▶ Answer questions based on texts
- ▶ Identify the sentiment of texts (e.g. reviews)

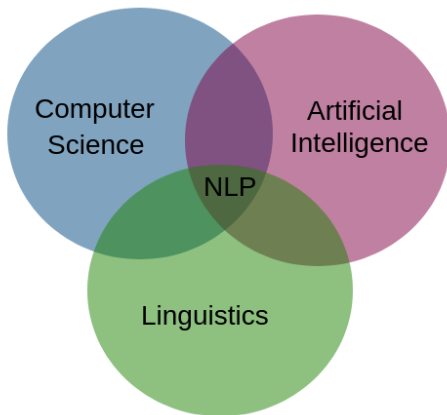
Natural Language Processing to the rescue

We are working on automatic methods to...



- ▶ Summarize multiple long texts
- ▶ Answer questions based on texts
- ▶ Identify the sentiment of texts (e.g. reviews)
- ▶ More...

What is Natural Language Processing (NLP)?

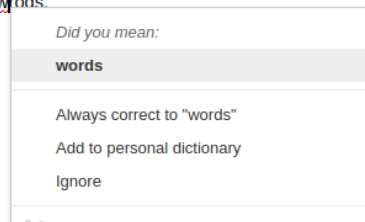


- ▶ **Goal:** for computers to “understand” and be able to communicate with people in natural languages (e.g. English)

NLP Applications are Everywhere

Spell Check

I don't make typos. I create new words.



NLP Applications are Everywhere

Grammar Correction

writing). But outside of those professions, there are few cases where using an expression that your readers may not understand would be better than writing it in **plain** English.

Review this sentence for vague words

The word plain is generic and can be overused. Consider using one of the following suggested replacements instead. Review the definitions of suggested words and select the replacement that fits the context.

Suggested replacements:

plain → [clear](#) ⓘ
[basic](#) ⓘ
[comprehensible](#) ⓘ
[intelligible](#) ⓘ
[understandable](#) ⓘ

Save time and ensure accuracy! [Have Grammarly correct all 8 issues for you.](#)

Was this card helpful? [👍 Yes](#) [👎 No](#)

[Next](#) →

Confusing modifiers

Other **1**

Punctuation within

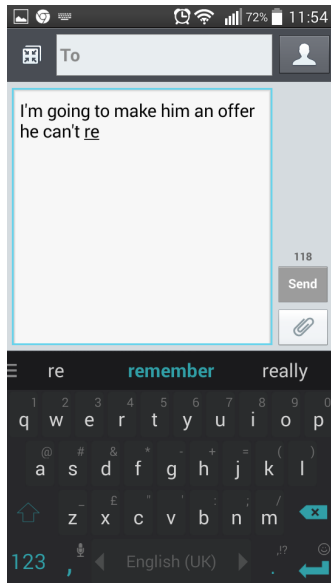
Vocabulary

Word usage issues, s
overuse of certain w

☰ Synonym

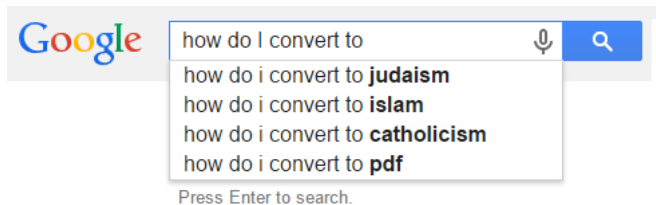
NLP Applications are Everywhere

Autocomplete



NLP Applications are Everywhere

Autocomplete



NLP Applications are Everywhere

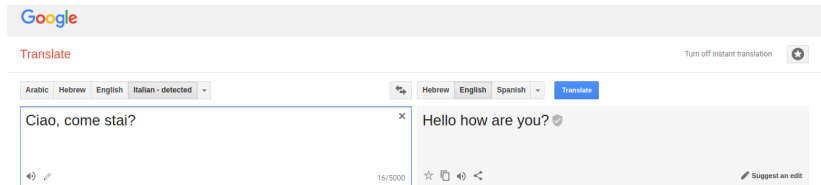
Spam Detection

[Delete all spam messages now](#)

<input type="checkbox"/> ☆	RMQ	Records Indicate You Now Qualify for a Reverse-Mortgage - Images N
<input type="checkbox"/> ☆	Remote Home Monitor (2)	Monitor your home from your phone, from anywhere in the world. - v
<input type="checkbox"/> ☆	Personal Cloud Service (2)	Secure access to your photos, music and files anytime, anywhere, at
<input type="checkbox"/> ☆	Bathroom Designs (2)	Fall back in love with your home by remodeling your bathroom. - Fal
<input type="checkbox"/> ☆	Career Upgrades (2)	Openings in the medical coding and billing field. - Upgrade to a Medi
<input type="checkbox"/> ☆	Asbestos exposure	Compensation for asbestos exposure. - Mesothelioma Resources How
<input type="checkbox"/> ☆	Newest electric cars (2)	Help the environment while saving money. - Save thousands on gas ev
<input type="checkbox"/> ☆	Train in nursing (2)	Consider a highly paid and flexible nursing career. - Have you been c
<input type="checkbox"/> ☆	Detect prostate troub. (2)	Detect discrete signs of prostate cancer. - We all get signs from our bo
<input type="checkbox"/> ☆	Advancement With an M. (2)	Twelve month MBA finished on the weekend - Vered going back to get

NLP Applications are Everywhere

Machine Translation



The screenshot displays the Google Translate web interface. At the top left is the Google logo. Below it, the word "Translate" is written in red. On the right side, there is a link "Turn off instant translation" and a circular icon with a plus sign. The main interface features a language selection bar with "Arabic", "Hebrew", "English", and "Italian - detected" (with a dropdown arrow). To the right of this bar are icons for voice input, a "Translate" button, and a language selection bar with "Hebrew", "English", and "Spanish" (with a dropdown arrow). The input text area on the left contains "Ciao, come stai?" and has a character count of "16/5000". The output text area on the right contains the translation "Hello how are you?" with a checkmark icon. Below the output text are icons for star, copy, voice, and share, along with a "Suggest an edit" link.

Google

Translate Turn off instant translation

Arabic Hebrew English Italian - detected

Hebrew English Spanish Translate

Ciao, come stai? 16/5000

Hello how are you? Suggest an edit

NLP Applications are Everywhere

Search Queries

Query

“Actors engaged in Scientology”

Results

Tom Cruise must ditch the vile cult of Scientology NOW before

www.dailymail.co.uk/.../PIERS-MORGAN-Tom-Cruise-ditch-vile-cult-Scie...

Apr 2, 2015 - PIERS MORGAN: I wanted to be Tom Cruise. As a fresh-faced, 21-year-old, I watched Top Gun a dozen times at my local movie theater in ...

John Travolta Says Scientology Is A Target Because It ...



www.huffingtonpost.com/.../john-travolta-scientology-target_n_710268...

Apr 20, 2015 - During an interview with "Good Morning America" Monday, John Travolta was asked why there is so much intrigue and interest surrounding the ...



NLP Applications are Everywhere

Question Answering


what is the weather tomorrow  

[All](#) [Images](#) [News](#) [Maps](#) [Videos](#) [More](#) [Settings](#) [Tools](#)

About 15,600,000 results (0.50 seconds)

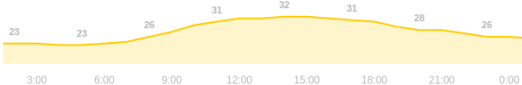
Bar-Ilan University, Ramat Gan

Tuesday
Mostly Sunny









 **33** °C | °F

Precipitation: 20%
Humidity: 31%
Wind: 18 km/h

[Temperature](#) [Precipitation](#) [Wind](#)



Time	Temperature (°C)
3:00	23
6:00	23
9:00	26
12:00	31
15:00	32
18:00	31
21:00	28
0:00	26

Day	Icon	High/Low (°C)
Mon		33° 22°
Tue		33° 22°
Wed		33° 21°
Thu		32° 21°
Fri		35° 21°
Sat		26° 18°
Sun		27° 19°
Mon		27° 19°

More on [weather.com](#) [Feedback](#)

NLP Applications are Everywhere

Targeted Ads



Vered Shwartz <vered@shwartz.com>

to Matt@shwartz.com

Hi Matt,

I'll be in **Seattle** on May 29-30 (before)

Thanks,
Vered

Vered, **Seattle** has some last-minute deals! [Trash](#)



Booking.com <email.campaign@sg.booking.com> [Unsubscribe](#)
to me

Tue, Apr 24, 9:10 AM (6 days ago)



This message has been deleted. [Restore message](#)

Booking.com

genius



city, region or property

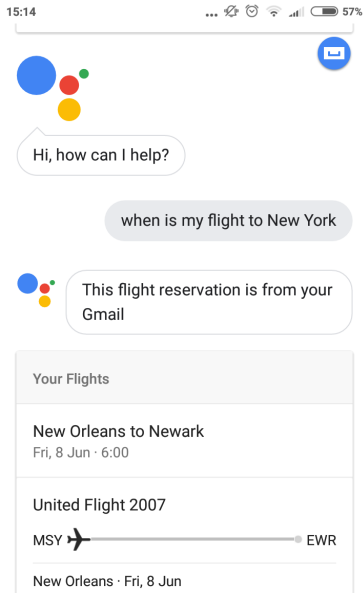
Search

genius

Your Genius status qualifies you for exclusive discounts on thousands of top properties.

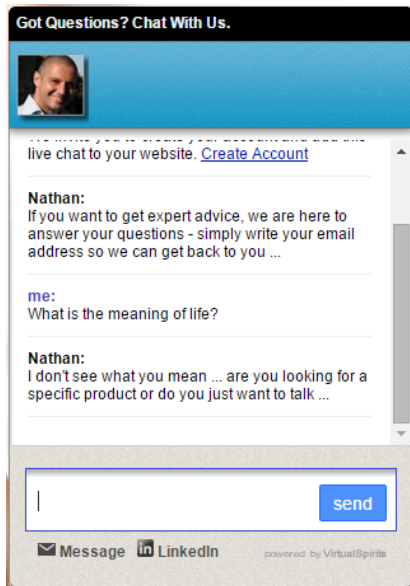
NLP Applications are Everywhere

Personal Assistants




NLP Applications are Everywhere

Chatbots



Got Questions? Chat With Us.





live chat to your website. [Create Account](#)

Nathan:
If you want to get expert advice, we are here to answer your questions - simply write your email address so we can get back to you ...

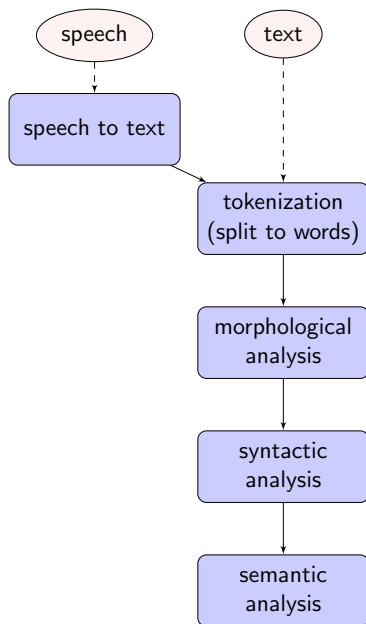
me:
What is the meaning of life?

Nathan:
I don't see what you mean ... are you looking for a specific product or do you just want to talk ...

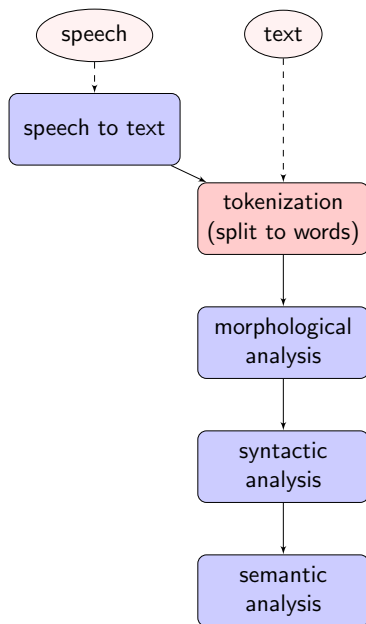
send

 Message  LinkedIn powered by VirtualSpirits

Text Analysis Tasks



Text Analysis Tasks



Text Analysis Tasks

Tokenization

- ▶ Split text into a sequence of tokens (\approx words)

Text Analysis Tasks

Tokenization

- ▶ Split text into a sequence of tokens (\approx words)
- ▶ Naive approach: split sentences by period, words by spaces

Text Analysis Tasks

Tokenization

- ▶ Split text into a sequence of tokens (\approx words)
- ▶ Naive approach: split sentences by period, words by spaces
- ▶ **How to tokenize this text?**

'Whose frisbee is this?' John asked, rather self-consciously.

'Oh, it's one of the boys' said the Sen.

Text Analysis Tasks

Tokenization

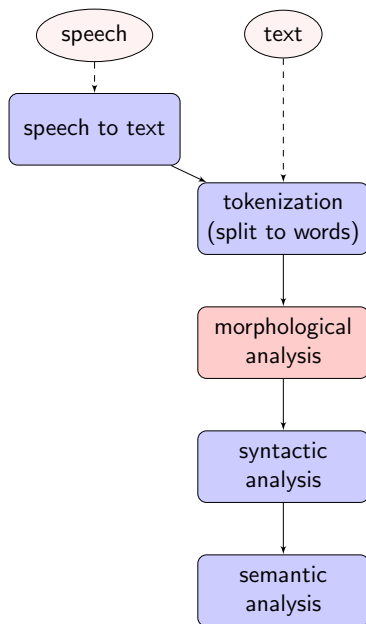
- ▶ Split text into a sequence of tokens (\approx words)
- ▶ Naive approach: split sentences by period, words by spaces
- ▶ **How to tokenize this text?**
*'Whose frisbee is this?' John asked, rather self-consciously.
'Oh, it's one of the boys' said the Sen.*
- ▶ **(Optional) answer:**

' Whose frisbee is this ?

' John asked , rather self-consciously.

' Oh , it 's one of the boys ' said the Sen.

Text Analysis Tasks



Text Analysis Tasks

Morphological Analysis

- ▶ Words are made from *morphemes*, smaller meaningful units

Text Analysis Tasks

Morphological Analysis

- ▶ Words are made from *morphemes*, smaller meaningful units
- ▶ Normally: base form + affixes

Text Analysis Tasks

Morphological Analysis

- ▶ Words are made from *morphemes*, smaller meaningful units
- ▶ Normally: base form + affixes
 - ▶ **Nouns** - plural form: dogs, suffixes, baby → babies

Text Analysis Tasks

Morphological Analysis

- ▶ Words are made from *morphemes*, smaller meaningful units
- ▶ Normally: base form + affixes
 - ▶ **Nouns** - plural form: dogs, suffixes, baby → babies
 - ▶ **Verbs** - tense: worked, working, person: works

Text Analysis Tasks

Morphological Analysis

- ▶ Words are made from *morphemes*, smaller meaningful units
- ▶ Normally: base form + affixes
 - ▶ **Nouns** - plural form: dogs, suffixes, baby → babies
 - ▶ **Verbs** - tense: worked, working, person: works
- ▶ Many irregularities... *“women and children begun running away as the wolves showed their teeth”*

Text Analysis Tasks

Morphological Analysis

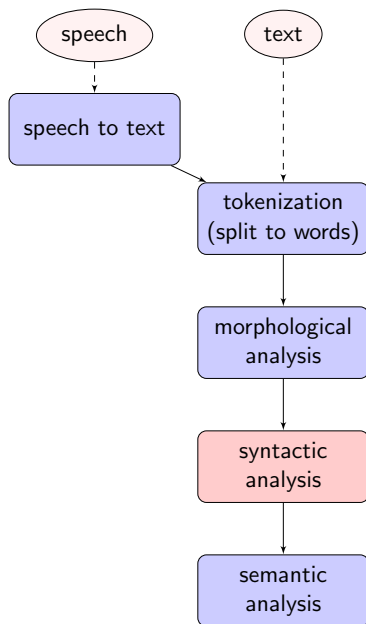
- ▶ Words are made from *morphemes*, smaller meaningful units
- ▶ Normally: base form + affixes
 - ▶ **Nouns** - plural form: dogs, suffixes, baby → babies
 - ▶ **Verbs** - tense: worked, working, person: works
- ▶ Many irregularities... *“women and children begun running away as the wolves showed their teeth”*
- ▶ Morphological analysis:
 - ▶ input: “am”, output: “be” + 1 PERSON + PRESENT

Text Analysis Tasks

Morphological Analysis

- ▶ Words are made from *morphemes*, smaller meaningful units
- ▶ Normally: base form + affixes
 - ▶ **Nouns** - plural form: dogs, suffixes, baby → babies
 - ▶ **Verbs** - tense: worked, working, person: works
- ▶ Many irregularities... *“women and children begun running away as the wolves showed their teeth”*
- ▶ Morphological analysis:
 - ▶ input: “am”, output: “be” + 1 PERSON + PRESENT
- ▶ Lemmatizer: reduce inflectional forms of a word to a common base form
e.g. children → child, running → run

Text Analysis Tasks



Text Analysis Tasks

Part of Speech Tagging

- ▶ Tags each word with its part of speech (POS): noun, verb, adjective, adverb, preposition, etc.

Part-of-Speech:

	<u>DT</u>	<u>JJ</u>	<u>NN</u>	<u>VBD</u>	<u>NN</u>	<u>NN</u>	<u>,</u>	<u>CC</u>	<u>RB</u>	<u>PRP</u>	<u>VBZ</u>	<u>VBG</u>	<u>TO</u>	<u>VB</u>
1	The	brown	dog	ate	dog	food,		and	now	he	is	going	to	sleep

Text Analysis Tasks

Part of Speech Tagging

- ▶ Tags each word with its part of speech (POS): noun, verb, adjective, adverb, preposition, etc.

Part-of-Speech:

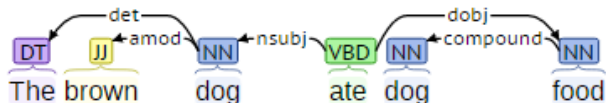
	DT	JJ	NN	VBD	NN	NN	CC	RB	PRP	VBZ	VBG	TO	VB
1	The	brown	dog	ate	dog	food,	and	now	he	is	going	to	sleep

- ▶ Surrounding words help deciding on the correct POS tag for ambiguous words:
I'm reading an interesting book ⇒ *book* = NOUN
I would like to book a flight ⇒ *book* = VERB

Text Analysis Tasks

Syntactic Parsing

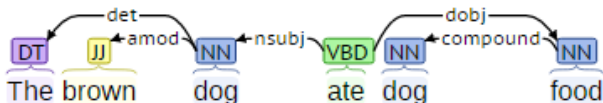
- Analyzes the syntactic structure of a sentence



Text Analysis Tasks

Syntactic Parsing

- ▶ Analyzes the syntactic structure of a sentence



- ▶ Let's look at some syntactic ambiguities!

Text Analysis Tasks

Syntactic Parsing

- ▶ *"They ate pizza with anchovies"*

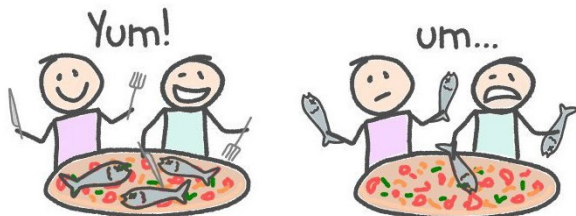


Creative Commons Attribution-NonCommercial 2.5
James Constable, 2010

Text Analysis Tasks

Syntactic Parsing

- ▶ *"They ate pizza with anchovies"*



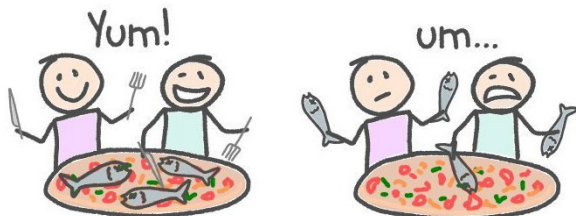
Creative Commons Attribution-NonCommercial 2.5
James Constable, 2010

- ▶ (1) They ate pizza, the pizza had anchovies on it

Text Analysis Tasks

Syntactic Parsing

- ▶ *"They ate pizza with anchovies"*



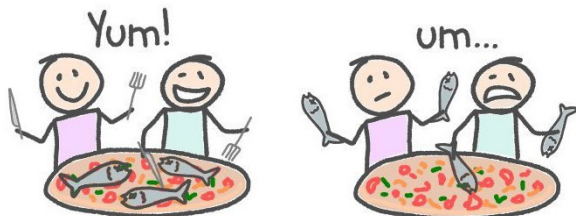
Creative Commons Attribution-NonCommercial 2.5
James Constable, 2010

- ▶ (1) They ate pizza, the pizza had anchovies on it
- ▶ (2) They ate pizza using anchovies instead of utensils

Text Analysis Tasks

Syntactic Parsing

- ▶ *"They ate pizza with anchovies"*

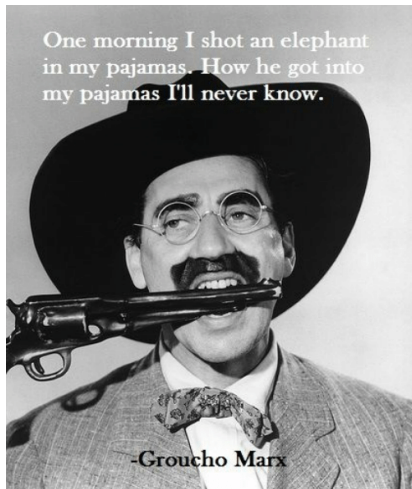


Creative Commons Attribution-NonCommercial 2.5
James Constable, 2010

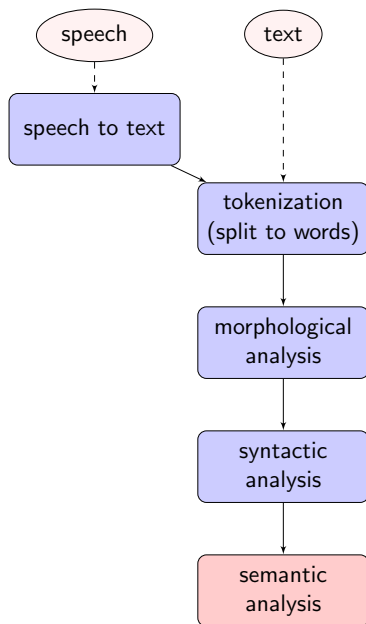
- ▶ (1) They ate pizza, the pizza had anchovies on it
 - ▶ (2) They ate pizza using anchovies instead of utensils
 - ▶ (3) The anchovies also ate pizza
-
- ▶ Each of the interpretations yields a different syntactic analysis

Text Analysis Tasks

Syntactic Parsing



Text Analysis Tasks

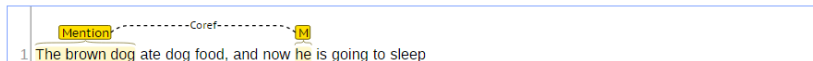


Text Analysis Tasks

Coreference Resolution

- Identify mentions referring to the same entity

Coreference:

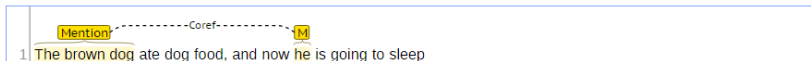


Text Analysis Tasks

Coreference Resolution

- ▶ Identify mentions referring to the same entity

Coreference:



- ▶ Considered a difficult task!

Text Analysis Tasks

Coreference Resolution



- ▶ *“I gave the monkeys the bananas because they were **hungry**”* ⇒
they = the monkeys

Text Analysis Tasks

Coreference Resolution



- ▶ *“I gave the monkeys the bananas because they were **hungry**”* ⇒
they = the monkeys
- ▶ *“I gave the monkeys the bananas because they were **ripe**”* ⇒
they = the bananas

Text Analysis Tasks

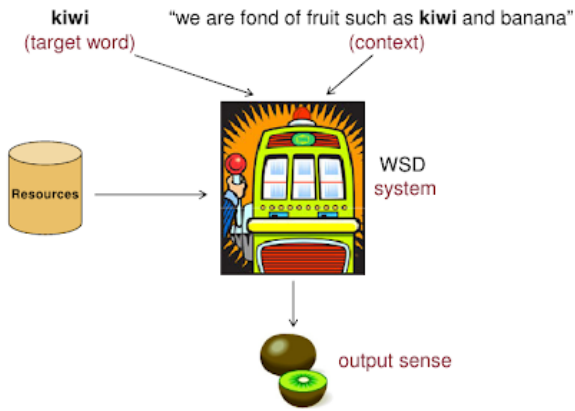
Word Sense Disambiguation

- ▶ What's the correct sense of a word in a given context?

Text Analysis Tasks

Word Sense Disambiguation

- ▶ What's the correct sense of a word in a given context?



from <http://naviglinlp.blogspot.co.il/>

Text Analysis Tasks

Named Entities

- ▶ **Named Entity Recognition:** recognize entities and their type

Person
John Doe worked in The Organization Ministry of Foreign Affairs Date last year.

Text Analysis Tasks

Named Entities

- ▶ **Named Entity Recognition:** recognize entities and their type

Person
John Doe worked in The Organization Ministry of Foreign Affairs Date last year.

- ▶ **Entity Linking:** linking entities to their Wikipedia pages

PERSON CARDINAL ORGANIZATION EVENT_COMMUNICATION
DATE PEOPLE DURATION ORDINAL

👤 🏠 🏈 KANSAS CITY, Mo -- There was no rational reason to expect Alex Smith to be in his current position.
👤 🕒 It was just a few years ago that he was a bust, a first-round pick of the 49ers who had failed to live up to expectations.
👤 🕒 🏈 His job had been snatched away by Colin Kaepernick and he had been shuttled off to Kansas City for a couple of draft picks, his career scuffling along but just barely.
👤 🏈 🏈 Chiefs offensive tackle Mitch Schwartz said, "He had a lot of adversity his first few years, had what, seven coordinators in seven years?"

Alex Smith
From Wikipedia, the free encyclopedia

For other people named Alex Smith, see Alex Smith (disambiguation).

Alexander Douglas Smith (born May 7, 1984) is an American football quarterback for the Kansas City Chiefs of the National Football League (NFL). He played college football at the University of Utah.



Kansas City Chiefs
From Wikipedia, the free encyclopedia

The Kansas City Chiefs are a professional American football team based in Kansas City, Missouri. The



Kansas City Chiefs
Current season

Established 1960; 58 years ago⁽¹⁾
First season 1960
Play in and headquartered in Arrowhead Stadium
Kansas City, Missouri

San Francisco 49ers
From Wikipedia, the free encyclopedia

This article needs additional citations for verification. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. (December 2016) (Learn how and when to remove this template message)



The San Francisco 49ers are a professional American football team located in the San Francisco Bay Area. They

San Francisco 49ers
Current season

Established 1946; 70 years ago⁽¹⁾
First season 1946

NLP is hard!

- ▶ Tokenization and POS tagging are almost 100% accurate today, but semantic tasks are far from that

NLP is hard!

- ▶ Tokenization and POS tagging are almost 100% accurate today, but semantic tasks are far from that
- ▶ Two major difficulties:

NLP is hard!

- ▶ Tokenization and POS tagging are almost 100% accurate today, but semantic tasks are far from that
- ▶ Two major difficulties:
 - ▶ **Ambiguity**: one text can have multiple meanings

NLP is hard!

- ▶ Tokenization and POS tagging are almost 100% accurate today, but semantic tasks are far from that
- ▶ Two major difficulties:
 - ▶ **Ambiguity**: one text can have multiple meanings
 - ▶ **Lexical variability**: the same meaning can be expressed with different words

21:41 75%

Hi, how can I help?

is it going to rain tomorrow

No, it won't rain tomorrow in Hod HaSharon. It'll be sunny, with a high of 34 and a low of 22.

Sunny on Wednesday Hod Hasharon • Precip: 20% **34°C**
High: 34° Low: 22°

23°	25°	27°	29°	31°	:
7	8	9	10	11	1

More on weather.com

Send daily What about this weekend?



21:41 75%

is rain expected tomorrow

You might want sunscreen

Sunny on Wednesday Hod Hasharon • Precip: 20% **34°C**
High: 34° Low: 22°

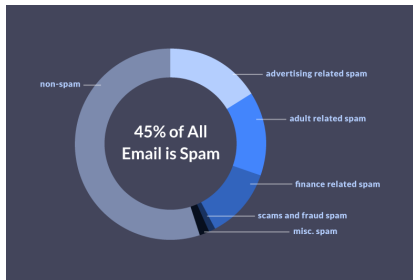
23°	25°	27°	29°	31°	:
7	8	9	10	11	1

More on weather.com

Send daily What about this weekend?

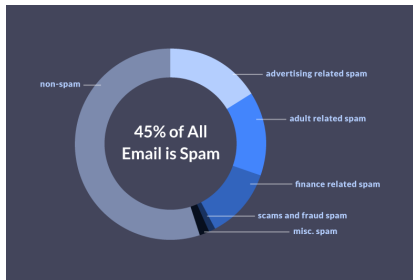


Example Application: Spam Detection



(used to be much worse... > 90%!)

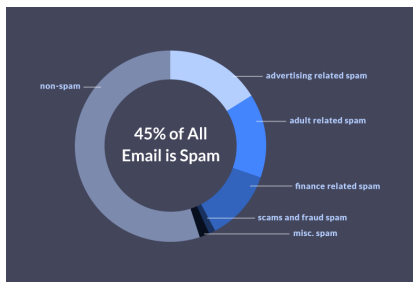
Example Application: Spam Detection



(used to be much worse... > 90%!)

- ▶ Automatically determine whether an email is spam or not

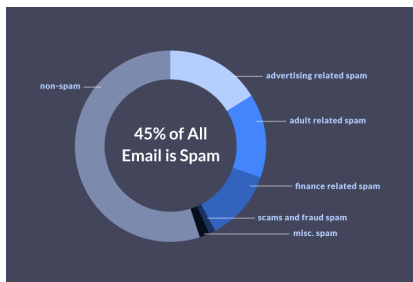
Example Application: Spam Detection



(used to be much worse... > 90%!)

- ▶ Automatically determine whether an email is spam or not
 - ▶ (and move spam messages to “spam” folder)

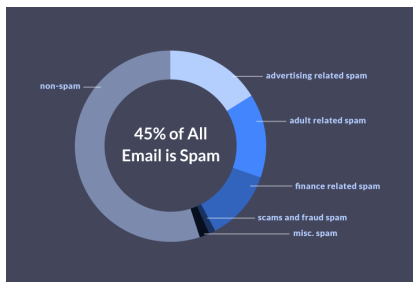
Example Application: Spam Detection



(used to be much worse... > 90%!)

- ▶ Automatically determine whether an email is spam or not
 - ▶ (and move spam messages to “spam” folder)
- ▶ Special case of *Text Classification*: given a text, automatically determine its topic

Example Application: Spam Detection



(used to be much worse... > 90%!)

- ▶ Automatically determine whether an email is spam or not
 - ▶ (and move spam messages to “spam” folder)
- ▶ Special case of *Text Classification*: given a text, automatically determine its topic
- ▶ How does it work?

Spam Detection

Let's think of characteristics of spam emails

- ▶ Unknown sender

Spam Detection

Let's think of characteristics of spam emails

- ▶ Unknown sender
- ▶ Spam triggering words:
 - ▶ *Earn extra cash*
 - ▶ *Earn \$*
 - ▶ *Free*
 - ▶ *Lose weight*
 - ▶ *Instant*
 - ▶ *Bonus*
 - ▶ ...

Spam Detection

Let's think of characteristics of spam emails

- ▶ Unknown sender
- ▶ Spam triggering words:
 - ▶ *Earn extra cash*
 - ▶ *Earn \$*
 - ▶ *Free*
 - ▶ *Lose weight*
 - ▶ *Instant*
 - ▶ *Bonus*
 - ▶ ...
- ▶ Naive idea: mark any email that contains these words as spam

Spam Detection

Let's think of characteristics of spam emails

- ▶ Unknown sender
- ▶ Spam triggering words:
 - ▶ *Earn extra cash*
 - ▶ *Earn \$*
 - ▶ *Free*
 - ▶ *Lose weight*
 - ▶ *Instant*
 - ▶ *Bonus*
 - ▶ ...
- ▶ Naive idea: mark any email that contains these words as spam
- ▶ **Problem:** inaccurate (will mark non-spam as spam and vice versa)

Spam Detection

Rule-based Approach

- ▶ Better idea: define rules, e.g. *“mark as spam if unknown sender and contains at least 2 spam triggering words”*

Spam Detection

Rule-based Approach

- ▶ Better idea: define rules, e.g. *“mark as spam if unknown sender and contains at least 2 spam triggering words”*
- ▶ More accurate: e.g. will not mark an email from your mother, with the word “instant” as spam :)

Spam Detection

Rule-based Approach

- ▶ Better idea: define rules, e.g. *“mark as spam if unknown sender and contains at least 2 spam triggering words”*
- ▶ More accurate: e.g. will not mark an email from your mother, with the word “instant” as spam :)
- ▶ **Problems:**
 - ▶ Finding the optimal rules is difficult

Spam Detection

Rule-based Approach

- ▶ Better idea: define rules, e.g. *“mark as spam if unknown sender and contains at least 2 spam triggering words”*
- ▶ More accurate: e.g. will not mark an email from your mother, with the word “instant” as spam :)
- ▶ **Problems:**
 - ▶ Finding the optimal rules is difficult
 - ▶ Not all triggering words were created equal

Spam Detection

Rule-based Approach

- ▶ Better idea: define rules, e.g. *“mark as spam if unknown sender and contains at least 2 spam triggering words”*
- ▶ More accurate: e.g. will not mark an email from your mother, with the word “instant” as spam :)
- ▶ **Problems:**
 - ▶ Finding the optimal rules is difficult
 - ▶ Not all triggering words were created equal
- ▶ **Solution:** Let the computer “learn” these rules alone!

Spam Detection

Supervised Learning

I have sent you this message earlier, but your failure to respond has prompted me to re-sending it once again. It is about my late client who lost his life in an automobile accident along with his wife and only child.

I assisted him in making a deposit worth **\$10.5M**. The Bank has therefore threatened to seize his account if an heir is not directly specified. You and my late client both share the same last name. With great respect, i want you to stand as an heir to the account so that his deposited funds can be released and transferred to you directly.

Kindly get back to my private email address for more update on this transaction
(richrdbernard65@gmail.com)

Best Regards

Barrister Richard Bernard.

- ▶ Let the computer learn a scoring function:

$$score = \dots + \alpha_{have} \cdot c(\text{have}) + \alpha_{sent} \cdot c(\text{sent}) + \dots + \alpha_{bernard} \cdot c(\text{bernard})$$

Spam Detection

Supervised Learning

I have sent you this message earlier, but your failure to respond has prompted me to re-sending it once again. It is about my late client who lost his life in an automobile accident along with his wife and only child.

I assisted him in making a deposit worth **\$10.5M**. The Bank has therefore threatened to seize his account if an heir is not directly specified. You and my late client both share the same last name. With great respect, i want you to stand as an heir to the account so that his deposited funds can be released and transferred to you directly.

Kindly get back to my private email address for more update on this transaction
(richrdbernard65@gmail.com)

Best Regards

Barrister Richard Bernard.

- ▶ Let the computer learn a scoring function:
 $score = \dots + \alpha_{have} \cdot c(\text{have}) + \alpha_{sent} \cdot c(\text{sent}) + \dots + \alpha_{bernard} \cdot c(\text{bernard})$
- ▶ Different weight α_j for each word, e.g. $\alpha_{cash} > \alpha_{document}$

Spam Detection

Supervised Learning

I have sent you this message earlier, but your failure to respond has prompted me to re-sending it once again. It is about my late client who lost his life in an automobile accident along with his wife and only child.

I assisted him in making a deposit worth **\$10.5M**. The Bank has therefore threatened to seize his account if an heir is not directly specified. You and my late client both share the same last name. With great respect, i want you to stand as an heir to the account so that his deposited funds can be released and transferred to you directly.

Kindly get back to my private email address for more update on this transaction
(richrdbernard65@gmail.com)

Best Regards

Barrister Richard Bernard.

- ▶ Let the computer learn a scoring function:
 $score = \dots + \alpha_{have} \cdot c(\text{have}) + \alpha_{sent} \cdot c(\text{sent}) + \dots + \alpha_{bernard} \cdot c(\text{bernard})$
- ▶ Different weight α_i for each word, e.g. $\alpha_{cash} > \alpha_{document}$
- ▶ Classify as spam if $score > threshold$ (learn threshold too!)

Spam Detection

Supervised Learning

- ▶ How does the computer learn the α weights?

Spam Detection

Supervised Learning

- ▶ How does the computer learn the α weights?
- ▶ **Supervised learning:** estimate a function (learn weights) using labeled examples

Spam Detection

Supervised Learning

- ▶ How does the computer learn the α weights?
- ▶ **Supervised learning:** estimate a function (learn weights) using labeled examples
- ▶ Take a lot of emails, manually mark them as spam/not spam

Spam Detection

Supervised Learning

- ▶ How does the computer learn the α weights?
- ▶ **Supervised learning:** estimate a function (learn weights) using labeled examples
- ▶ Take a lot of emails, manually mark them as spam/not spam
- ▶ The computer learns a function (weights) that best predicts spam/not spam for the **known** emails

Spam Detection

Supervised Learning

- ▶ How does the computer learn the α weights?
- ▶ **Supervised learning:** estimate a function (learn weights) using labeled examples
- ▶ Take a lot of emails, manually mark them as spam/not spam
- ▶ The computer learns a function (weights) that best predicts spam/not spam for the **known** emails
- ▶ If we have enough examples, it would also work well on new emails

Spam Detection

Features

- ▶ We used *bag-of-words* as features for classification :
{ I, have, sent, you, ... }

Spam Detection

Features

- ▶ We used *bag-of-words* as features for classification :
{ I, have, sent, you, ... }
- ▶ If we have enough spam examples that contain the word “urgent”, α_{urgent} will be high

Spam Detection

Features

- ▶ We used *bag-of-words* as features for classification :
{ I, have, sent, you, ... }
- ▶ If we have enough spam examples that contain the word “urgent”, α_{urgent} will be high
- ▶ What about similar words like “immediate” or “instant”?

Spam Detection

Features

- ▶ We used *bag-of-words* as features for classification :
{ I, have, sent, you, ... }
- ▶ If we have enough spam examples that contain the word “urgent”, α_{urgent} will be high
- ▶ What about similar words like “immediate” or “instant”?
- ▶ We need to find a way to let the computer know about semantically-similar words

Word Representation

One-hot Vectors

- ▶ How do we represent all the words in the computer?

Word Representation

One-hot Vectors

- ▶ How do we represent all the words in the computer?
- ▶ **Simplest:** we have a dictionary, and each word has an index, e.g. $index(\text{urgent}) = 316$, $index(\text{instant}) = 12418$

Word Representation

One-hot Vectors

- ▶ How do we represent all the words in the computer?
- ▶ **Simplest:** we have a dictionary, and each word has an index, e.g. $index(\text{urgent}) = 316$, $index(\text{instant}) = 12418$
- ▶ You can think of the word with index i as a vector (array of numbers) with zeros and one entry with 1 in the i th index - “one-hot vector”:

urgent

0	0	...	1	0	...	0	0	...	0
---	---	-----	---	---	-----	---	---	-----	---

↑
316

instant

0	0	...	0	0	...	1	0	...	0
---	---	-----	---	---	-----	---	---	-----	---

↑
12418

Spam Detection

Bag-of-words with One-hot Vectors

- ▶ A vector representing the entire email: sum of one-hot vectors of the words in the email:

I	0	0	...	1	0	...	0	0	...	0
have	0	1	...	0	0	...	0	0	...	0
sent	0	0	...	0	0	...	1	0	...	0
+				
bernard	0	0	...	0	1	...	0	0	...	0
=										
feature vector	0	4	...	2	1	...	1	0	...	0

Spam Detection

Bag-of-words with One-hot Vectors

- ▶ A vector representing the entire email: sum of one-hot vectors of the words in the email:

$$\begin{array}{r} \text{I} \\ \text{have} \\ \text{sent} \\ + \dots \\ \text{bernard} \\ = \\ \text{feature vector} \end{array} \begin{array}{|c|c|c|c|c|c|c|c|c|c|} \hline 0 & 0 & \dots & 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ \hline 0 & 1 & \dots & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ \hline 0 & 0 & \dots & 0 & 0 & \dots & 1 & 0 & \dots & 0 \\ \hline \dots & & & & & & & & & \\ \hline 0 & 0 & \dots & 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \hline \hline 0 & 4 & \dots & 2 & 1 & \dots & 1 & 0 & \dots & 0 \\ \hline \end{array}$$

- ▶ **Problem:** Emails with similar words (e.g. *deliver* instead of *send*, *urgent* instead of *instant*) have very different feature vectors!

Word Representation

Distributional Word Vectors

- ▶ Can we have similar vectors for semantically-similar words?

Word Representation

Distributional Word Vectors

- ▶ Can we have similar vectors for semantically-similar words?
- ▶ “*You shall know a word by the company it keeps*”
(John Rupert Firth, 1957)

Word Representation

Distributional Word Vectors

- ▶ Can we have similar vectors for semantically-similar words?
- ▶ “*You shall know a word by the company it keeps*”
(John Rupert Firth, 1957)

elevator	0	0	...	0.16	0	...	0.49	0	...	0
lift	0	0	...	0.15	0	...	0.51	0	...	0

↑ ↑
up stairs

Word Representation

Distributional Word Vectors

- ▶ Can we have similar vectors for semantically-similar words?
- ▶ *"You shall know a word by the company it keeps"*
(John Rupert Firth, 1957)

elevator	0	0	...	0.16	0	...	0.49	0	...	0
lift	0	0	...	0.15	0	...	0.51	0	...	0

↑ ↑
up **stairs**

- ▶ Now semantically-similar words have similar word vectors!

Spam Detection

Bag-of-words with Distributional Word Vectors

- ▶ Again, we sum up all the vectors:

I	0	0	...	0.12	0.03	0.04	0	...	0
have	0	0.22	...	0	0	0	0	...	0
sent	0	0.43	...	0	0.1	0.25	0	...	0
+							
bernard	0	0	...	0	0.67	0	0	...	0
=													
FV	0	0.65	...	0.12	0.71	0.29	0	...	0

Spam Detection

Bag-of-words with Distributional Word Vectors

- ▶ Again, we sum up all the vectors:

I	0	0	...	0.12	0.03	0.04	0	...	0
have	0	0.22	...	0	0	0	0	...	0
sent	0	0.43	...	0	0.1	0.25	0	...	0
+							
bernard	0	0	...	0	0.67	0	0	...	0
=													
FV	0	0.65	...	0.12	0.71	0.29	0	...	0

- ▶ We can now replace a word (e.g. *sent*) with a similar word (e.g. *delivered*) and get a similar feature vector \Rightarrow same classification for similar emails!

Word Embeddings

- ▶ [A more recent type of distributional vectors]
- ▶ Find most similar words:

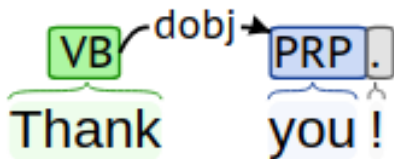
Nearest words

Given a word, this demo shows a list of other words that are similar to it, i.e. nearby in the vector space.

Case sensitive: **Top N:**

Manhattan
NY
Brooklyn
Long_Island
NYC
upstate
midtown_Manhattan
New_Jersey
Greenwich_Village
Bronx

See more here: http://bionlp-www.utu.fi/wv_demo/



Additional Resources

- ▶ Books:
 - ▶ Chris Manning and Hinrich Schütze, Foundations of Statistical Natural Language Processing, MIT Press. Cambridge, MA: May 1999.
 - ▶ Dan Jurafsky and James H. Martin, Speech and Language Processing. Second Edition. Pearson Education, 2014.
- ▶ Resources from NACLO - North American Computational Linguistics Olympiad
<http://nacloweb.org/resources.php>
- ▶ My blog: <http://veredshwartz.blogspot.co.il>