

Natural Language Processing

Lec 1: Introduction

Some slides in this lecture are inspired by Dr. Kai-wei Chang UCL, Stanford course CS124.



Content

Work Load



Gregory R. Hancock
@GregoryRHancock

dear students...



Tentative Grading

Assessment	Percentage
Year Work (Assignments+ Quizzes+ Project)	30%
Midterm Exam	20%
Final Exam	50%

Cheating Policy

- No, you can not cheat. If you have any concern, please ask.
- **Zero** for any assignment or the mid-term if discovered cheating.
- If discovered cheating in more than one assignment, **25% will be deducted from the full mark.**
- Always cite your references and clearly state what are your contributions.

Why is this course?
What is NLP?



Why is this course?

Artificial Intelligence is one of the most interesting fields of research today, and language is the most compelling manifestation of intelligence.

Phil Blunson, Google Deep Mind, University of Oxford



What is NLP?

- **Wiki: Natural language processing (NLP)** is a field of computer science, artificial intelligence, and computational linguistics concerned with the interactions between computers and human (**natural**) languages.



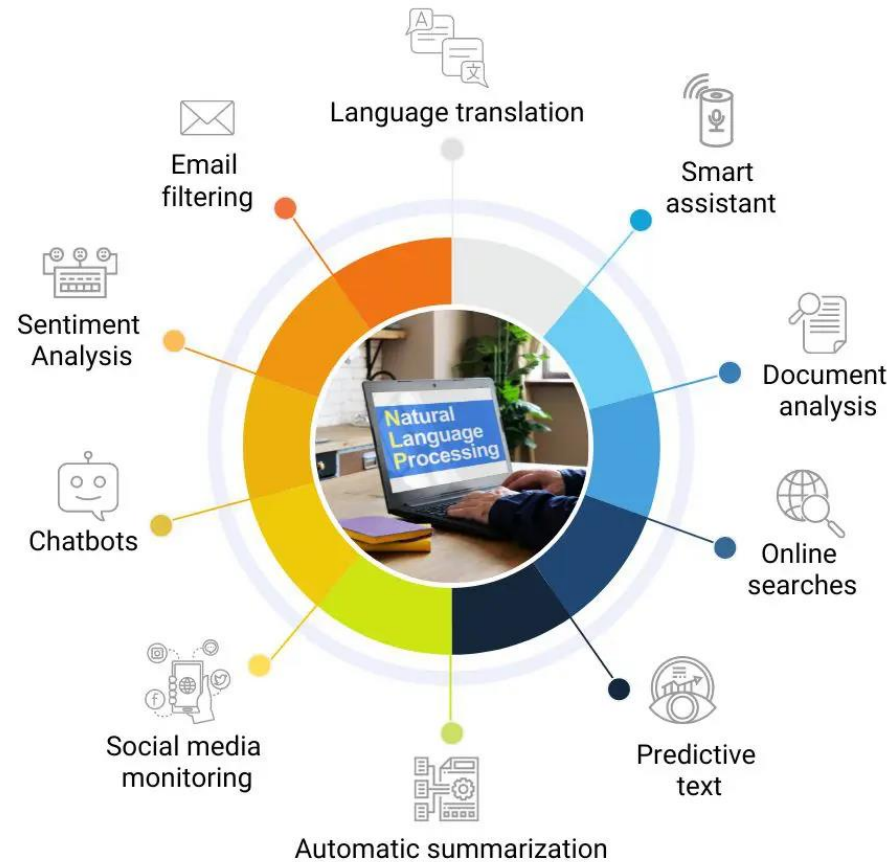
Go Beyond the Text

- Identify the **structure** and **meaning** of **words**, **sentences**, **texts** and **conversations**.
- **Deep** understanding of **broad** language



NLP is everywhere

Applications of Natural Language Processing



Machine Translation

Haaretz הארץ
חיום: נתניהו שוב ינסה להעביר את ד"ח טרכטנברג בממשלה
www.haaretz.co.il

אחרי הפאריסה בשבוע שעבר, בלשכת רה"מ מסרבים להתחייב שתהיה הצבעה בתום הדיון, בישראל ביתנו, ש"ס והעצמאות דורשים ריכוך או השמטה של חלק מההמלצות

Like · Comment · Translate · Share · Yesterday at 06:00

9 people like this.

View 1 share

שוקי זקס With full pay the price for hzhitho vakombinot of the Prime Minister. Not find it appropriate to action lhabrat the health system. Endangered animals, people and caused the suffering that must drive with ... city. Foundation of Minneapolis had already thrown him. Europe also. This man destroys the country with the citizens in its path. Contempt you Binyamin Netanyahu. Netanyahu was not committed to voting the Trachtenberg managed to impose a "partnership hkoalition terror the Prime Minister, who is pledging to bring voting day the conclusions report Trachtenberg. hourly opposes. Israel Beiteinu-independence morning ministerial session will vihalto only it-political system, news

עמי וקנין Or acknowledging the report moves or the chips will go

Dalya Gumis Stop vote once to

Yuval Gilor Tired already go home

Haaretz הארץ
חיום: נתניהו שוב ינסה להעביר את ד"ח טרכטנברג בממשלה
www.haaretz.co.il

אחרי הפאריסה בשבוע שעבר, בלשכת רה"מ מסרבים להתחייב שתהיה הצבעה בתום הדיון, בישראל ביתנו, ש"ס והעצמאות דורשים ריכוך או השמטה של חלק מההמלצות

Like · Comment · Original · Share · Yesterday at 06:00

9 people like this.

View 1 share

שוקי זקס With full pay the price for hzhitho vakombinot of the Prime Minister. Not find it appropriate to action lhabrat the health system. Endangered animals, people and caused the suffering that must drive with ... city. Foundation of Minneapolis had already thrown him. Europe also. This man destroys the country with the citizens in its path. Contempt you Binyamin Netanyahu. Netanyahu was not committed to voting the Trachtenberg managed to impose a "partnership hkoalition terror the Prime Minister, who is pledging to bring voting day the conclusions report Trachtenberg. hourly opposes. Israel Beiteinu-independence morning ministerial session will vihalto only it-political system, news

עמי וקנין Or acknowledging the report moves or the chips will go

Dalya Gumis Stop vote once to

Yuval Gilor Tired already go home

Detect language **Spanish** English Arabic ▼

Hola chicos, buenos dias

[Look up details](#)

24 / 5,000


↔ **English** Arabic Spanish ▼

Hello guys, good morning

[Look up details](#)

Send feedback

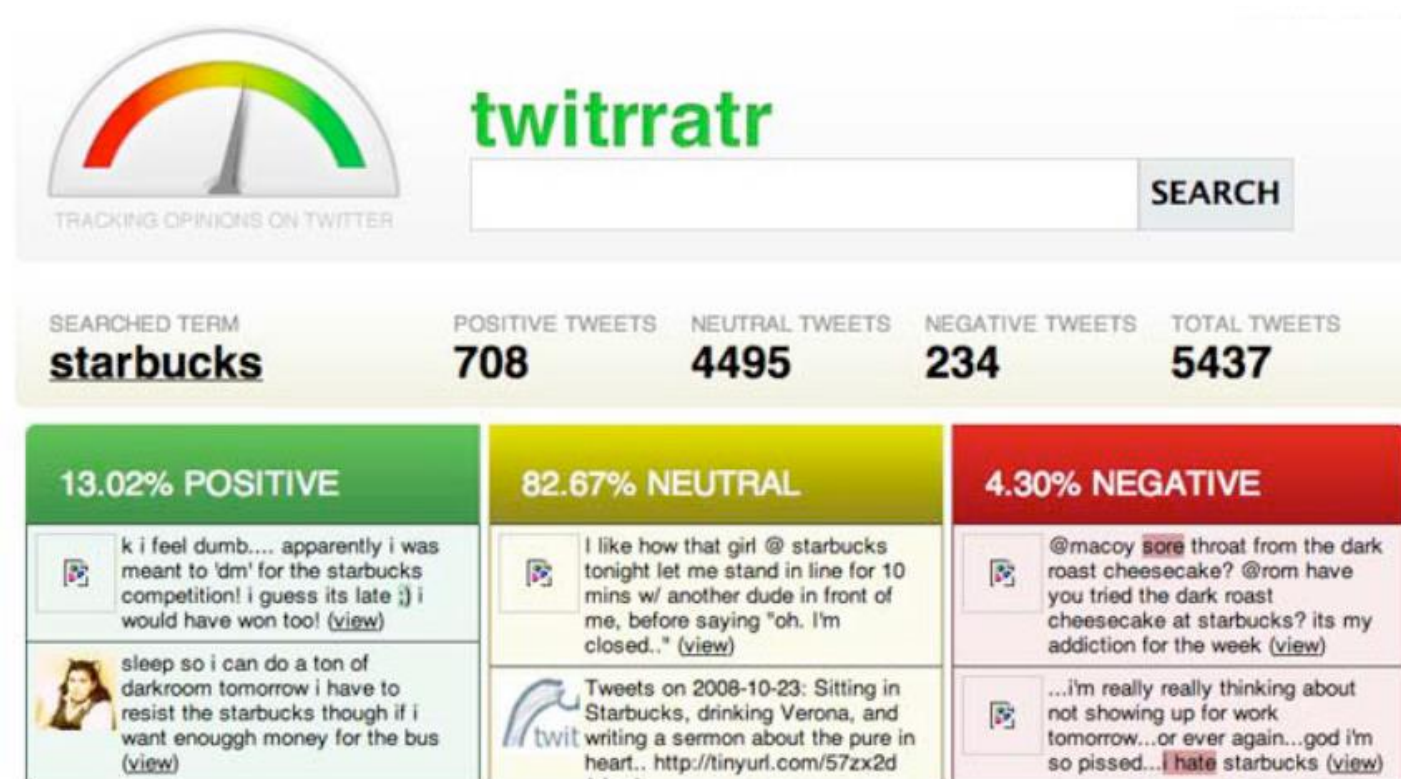
Language model

san fe 

san fernando valley
san fernando mission
san fernando
san fernando high school
san felipe
san fernando weather
san felipe mexico
san fernando valley news
san fernando court
san fernando brewery

[Google Search](#) [I'm Feeling Lucky](#)

Sentiment/Opinion Analysis



Text Classification



1-21 of 21

Primary Social 1 new Promotions 2 new Updates 1 new

<input type="checkbox"/>	<input type="checkbox"/>	James, me (2)	Hiking	Hiking trip on Saturday - Yay - so glad you can join. We should leave from I	3:14 pm
<input type="checkbox"/>	<input type="checkbox"/>	Hannah Cho	Thank you	- Keri - so good that you and Steve were able to come over. Thank you :	3:05 pm
<input type="checkbox"/>	<input type="checkbox"/>	Jan Birdson	School	Upcoming school conference dates - Hello everyone. A few people have	12:26 pm

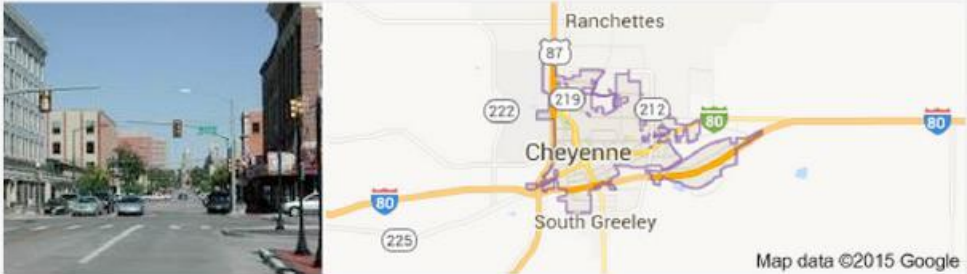
Question and Answering

What's the capital of Wyoming?

Web Maps Shopping Images News More Search tools

About 984,000 results (0.54 seconds)


Wyoming / Capital



Cheyenne

Information Extraction

Event: Curriculum mtg
Date: Jan-16-2026
Start: 10:00am
End: 11:30am
Where: Gates 159

- Subject: **curriculum meeting**
- Date: January 15, 2026
 - To: Dan Jurafsky
- Hi Dan, we've now scheduled the curriculum meeting.
- It will be in Gate  30.
- -Chris



Challenges

Challenges: Ambiguity

- Ambiguity at multiple levels
- Word senses: **bank** (finance or river ?)
- Part of speech: **chair** (noun or verb ?)
- Syntactic structure: **I can see a man with a telescope**



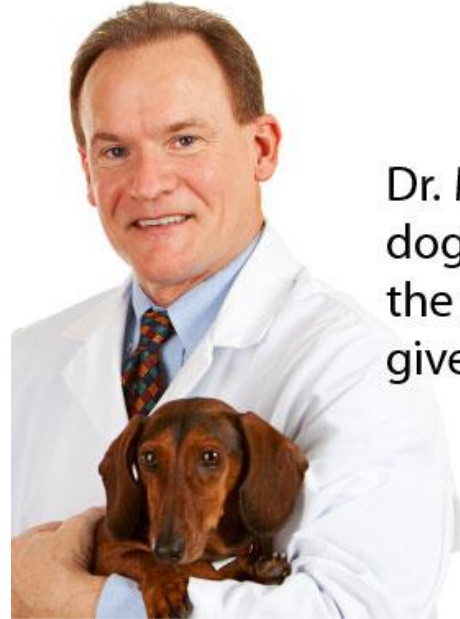
Challenges: Ambiguity

- Word sense / meaning ambiguity



Challenges: Ambiguity

- Pronoun reference ambiguity



Dr. Macklin often brings his dog Champion to visit with the patients. **He** just loves to give big, wet, sloppy kisses!

Challenges: Language is not Static

- Language grows and changes

LOL	
G2G	
BFN	
B4N	
Idk	
FWIW	
LUWAMH	

Challenges: Language is Compositional



Challenges: Languages are very diverse

- Different languages
- Different gender expressions
- Different levels of grammar (High-grammatical, Neutral, low-grammatical)



Challenges: Scale

- Examples:
- Bible (King James version): ~700K
- Penn Tree bank ~1M from Wall street journal
- Newswire collection: 500M+
- Wikipedia: 2.9 billion word (English)
- Web: several billions of words



Enough?!

- Enough! OK...any questions?!

