

CE/CZ4045 Natural Language Processing

Lecture 00 Introduction



Your Instructor

- Dr. Sun Aixin
 - Email: axsun@ntu.edu.sg
 - Email subject: **[CE/CZ4045]**
 - Office: N4-02C-102 Phone: 6790-5139
 - Homepage: www.ntu.edu.sg/home/axsun
 - Research Interests: Text Mining, Information Retrieval, Social Computing ...



Outline for today

Natural Language Processing (NLP)

- Course Expectation
 - Evaluation
- What is NLP?/ Why take NLP course?
 - Example application

Can

Can ah?	Can you or can't you?	Can hor	You are sure then...
Can lah	Yes.	Can meh?	Are you certain?
Can leh	Yes. I think so.	Can bo?	Can or not?
Can lor	Yes. Of course.	Can can	Confirm
Can hah?	Are you sure?	Can liao	Already can / Done

Semantics

ANGMOHDAN.COM

One Sentence about NLP

- Teaches key theory and methods for NLP:
 - word-level analysis, parsing, semantics, etc.
 - Learn techniques which can be used in **practical**, robust systems that can (partly) understand human language
- This is **not** a language course
 - Computational methods of processing natural languages
 - But, you are expected to have knowledge of (basic) English grammar

Preparation

- Pre-requisites
 - Basic understanding on English grammar,
 - e.g., verb, noun phrase, preposition
 - Basic algorithm and data structure analysis,
 - e.g., dynamic programming
 - Basic probability concepts,
 - e.g., conditional probability

$$P(B|A) = \frac{P(A, B)}{P(A)}$$

My house is on top of that hill.

Possessive pronoun, noun, verb, preposition, noun,
preposition, determiner, noun

Noun phrase, verb phrase, prepositional phrase

- **Decent** programming skills

Divide & Conquer	Dynamic Programming
Partitions a problem into independent smaller sub-problems	Partitions a problem into overlapping sub-problems
Doesn't store solutions of sub-problems. (Identical sub-problems may arise – results in the same computations are performed repeatedly)	Stores solutions of sub-problems: thus avoids calculations of same quantity twice.
Top down algorithms: which logically progresses from the initial instance down to the smallest sub-instances via intermediate sub-instances.	Bottom up algorithms: in which the smallest sub-problems are explicitly solved first and the results of these used to construct solutions to progressively larger sub-instances.

Preparation (Cont'd)

Machine learning?

- Machine learning knowledge can be **helpful** for assignment and some parts of lecture
- Not everyone has the same skills
 - Assumes some ability to learn missing knowledge

What computation?

- **Lots** of statistics!
- Some rules based on linguistic theory



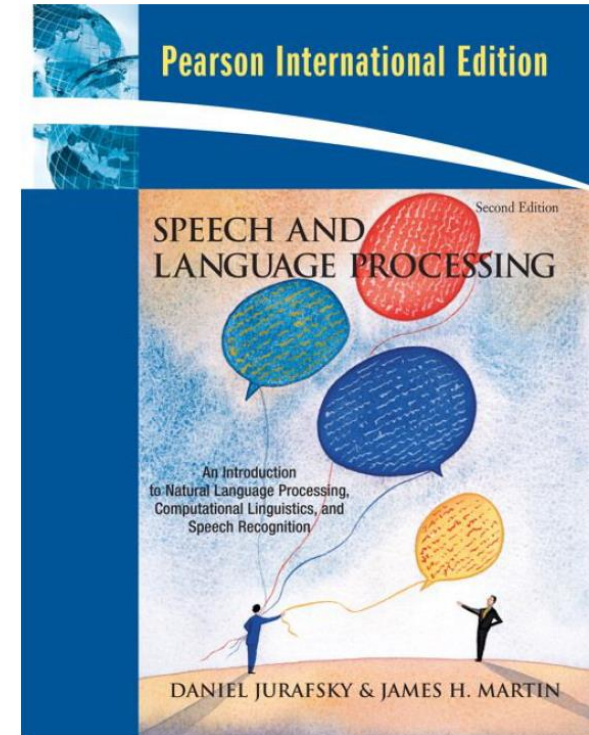
Course Evaluation

- Evaluation Objective
 - Spread evaluation over the whole course, not just one exam or one report
- Main Components
 - 35% assignment (group assignment)
 - 15% mid-term quiz
 - 50% final exam



Textbook

- SPEECH and LANGUAGE PROCESSING
 - [Daniel Jurafsky](#) and [James H. Martin](#), 2nd edition, Prentice-Hall 2009
 - A copy is kept in Library, in the reserve section
 - Draft of the 3rd edition:
<https://web.stanford.edu/~jurafsky/slp3/>
- Recommended reading
 - Steven Bird, Ewan Klein, and Edward Loper. Natural Language Processing with Python. 1st edition, O'Reilly Media; 2009.
(<http://www.nltk.org/book/>)



Schedule (tentative)

Week	Lecture topic	Tutorial topic
1	Introduction, Regular Expression (Chap 1, 2)	-
2	Word-level Analysis (Chap 3)	-
3	N-grams (Chap 4)	Regular Expression
4	N-grams (Chap 4)	Word-level Analysis
5	POS Tagging (Chap 5)	N-grams
6	Hidden Markov Model (Chap 6)	POS, HMM
7	(Machine Translation) (Chap 25)	
8	Formal Grammar (Chap 12)	POS, HMM
9	Syntactic Parsing (Chap 13)	Formal Grammar
10	Statistical Parsing (Chap 14)	Syntactic Parsing
11	NLP Applications (Chap 22 – 24)	Statistical Parsing
12	Semantic Analysis (Chap 17 – 21)	NLP Applications



Expectations

- You are final year students;
 - You are **willing to learn NLP**
- You are expected to participate.
- You are expected to
 - **Read** lecture slides for reference only
 - **Read** J&M book and other references
 - **Enjoy** assignment and programming!



Goals of the field of NLP

- Computers would be a lot more useful if they could
 - handle our email, do our library research, chat to us...
- But someone has to work on the hard problems!
 - How can we tell computers about language?
 - Or help them learn it as kids do?
- In this course we seek to identify many of **the open research problems** in natural language processing

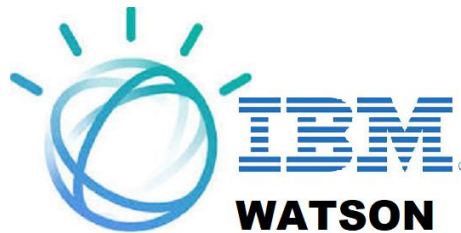


What/where is NLP?

- Goals can be very far reaching ...
 - True text understanding
 - Real-time participation in spoken dialogs
- Or very down-to-earth ... (the Web, business documents)
 - Finding the price of products on the web
 - Sentiment detection about products or stocks
 - Extracting facts or relations from documents
- These days, the latter predominate (as NLP becomes increasingly practical, it is increasingly engineering oriented)



Commercial World: **Lots** of exciting stuff going on in industry



Chinese (Simplified) ▼	English ▼
知之为知之，不知为不知，是知也。 Edit	Know as know , I do not know as I do not know , that is knowledge .
English ▼	Chinese (Simplified) ▼
Know as know , I do not know as I do not know , that is knowledge . Edit	知道的知道，我不知道，因为我不知道，那就是知识。 Zhīdào de zhīdào, wǒ bù zhīdào, yīnwèi wǒ bù zhīdào, nà jiùshì zhīshì.
Know, admit what you don't know, is know.	
Open in Google Translate	

“It is wise to hold what you know and admit what you don't know.”
– Baidu Zhidao




Example of down-to-earth Applications

- Some deployed applications
 - Machine translation: Chinese < == > English
 - Question answering: Yahoo! Answer, Baidu Zhidao
 - Information extraction: Extracting product information from the Web
 - Text analytics: Sentiment Analysis
- Example <https://alchemy-language-demo.mybluemix.net/>

IBM Watson Developer Cloud

Services



AlchemyLanguage

AlchemyLanguage is a collection of APIs that offer text analysis through natural language processing. This set of APIs can analyze text to help you understand its concepts, entities, keywords, sentiment, and more. Additionally, you can create a custom model for some APIs to get specific results that are tailored to your domain.

Google Translate



Killing Palestinians and wounding nine in the raids Sector

Nine Palestinians were wounded among civilians in an Israeli air raid in the neighborhood result in the Gaza Strip. This comes immediately after the killing of two prominent Al-Aqsa Martyrs Brigades in the Israeli occupying forces carried out air and infantry forces in the Balata camp in the West Bank.



Bashir meets Fraser, the Security Council will not impose forces Darfur

Is scheduled to meet with Sudanese President Omar al-Bashir Jenday Fraser Assistant Minister for Foreign Affairs of the American attempt to persuade officials in Khartoum, Sudanese Darfur deployment of the nationalities. For his part, US Ambassador to the United Nations that it has no intention of the Security Council to impose its forces in the province.



Rmsfield and Cheney insist on keeping the American forces in Iraq

Called American Defense Minister Donald Rmsfield Americans to show patience on Iraq. I take Vice President Dick Cheney calls Democrats withdrawal of American forces from Iraq link and the possibility of early withdrawal of attacks inside the United States.



Killing civilians and wounding officer suicide attack in Afghanistan

The international force to help establish security (ISAF) killed civilians and the wounding of an officer in an attack against Afghan forces convoy south Atlantic Afghanistan. In the capital Kabul, a hand grenade exploded at the passage of manufacture French patrol was not reported injuries or damage.



Web Q/A

- Get answers directly
 - Without the need of clicking, Recommend related questions, Retrieve relevant information

Google what is the population of singapore

All Maps Images News Videos More Search tools

About 50,600,000 results (0.78 seconds)

Singapore / Population

5.399 million (2013)

Year	Hong Kong (2013)	Singapore (2013)	Kuala Lumpur
1960	~2.5M	~1.5M	~0.5M
1970	~3.5M	~2.0M	~0.7M
1980	~4.5M	~2.5M	~0.9M
1990	~5.5M	~3.0M	~1.1M
2000	~6.5M	~4.0M	~1.3M
2010	~7.0M	~5.0M	~1.5M
2013	7.188 million	5.399 million	-

Explore more

Sources include: World Bank, UNdata

Feedback

People also ask

Is Singapore population increasing?

[Singapore population 2016 | Current population of Singapore](#)
[countrymeters.info/en/Singapore](#)

As of 1 January 2016, the population of Singapore was estimated to be 5 664 322 people. ... The sex ratio of the total population was 1.017 (1 017 males per 1 000 females) which is lower than global sex ratio. ... During 2016 Singapore population is projected to increased by 111 304 people ...

Singapore

Country in Asia

Singapore, an island city-state off southern Malaysia, is a global financial centre with a tropical climate and multicultural population. In circa-1820 Chinatown stands the red-and-gold Buddha's Tooth Relic Temple, Little India offers colorful souvenirs and Arab Street is lined with fabric shops. Singapore is also known for eclectic street fare, served in hawker centres such as Tiong Bahru and Maxwell Road.


Related statistics

GDP per capita	55,182.48 USD (2013)
Gross domestic product	297.9 billion USD (2013)
Life expectancy	82.14 years (2012)

Population elsewhere

Malaysia	29.72 million (2013)
Thailand	67.01 million (2013)
Australia	23.13 million (2013)

Another example in Web search

Google 



[All](#) [Images](#) [Videos](#) [News](#) [Maps](#) [More ▾](#) [Search tools](#)

About 32,500 results (0.48 seconds)

Updated 1d 3h ago

Singapore Airlines Flight 800

SCHEDULED

SIN   **PEK**

Singapore · Thu, 4 Aug			Beijing · Thu, 4 Aug		
Scheduled departure	Terminal	Gate	Scheduled arrival	Terminal	Gate
1:10 AM	3	-	7:25 AM	3	-

Community based QnA

YAHOO! ANSWERS

Q Search

Search W

HOME

BROWSE CATEGORIES

MY ACTIVITY

ABOUT

Ask

What would you like to ask?

Continue

Answer

Discover

Home > Search Results for "What is the cost of living in Singapore?"

Search

What is the cost of living in Singapore?

Search Yahoo! Answers

Advanced Search

stackoverflow

Questions

Jobs

Documentation Beta

Tags

Users

Badges

Ask Question

Refine search

Keyword match in

All Q&A

Category

All Categories

Location

All English questions

Question Status

All Questions

Number of Answers

Any Number of Answers

Date Submitted

Anytime

Sort by: Relevance | Newest | Most Ans

Cost of Living Calculator

Side-by-Side City Comparison of Categories

Bankrate.com/CostOfLiving

Singapore Job Search

Post Your Resume 100% Free. And Get Hin

SingaporeJobs77.com

! singapore living cost for family

...relocate to singapore from india in

my wife and small baby boy. my job pl

Bangkok Bank Building Singapore 06

the exp for living. and what should be

☆ In Singapore - Asked by deepak - 2 a

What is the cost of living in Si

salary for 9yrs IT experience?

... for a job opportunity in Singapore.

there and reasonable package that I...

2 ☆ In Singapore - Asked by girlishere -

All Questions

newest

407 featured

frequent

votes

active

unanswered

6 votes

+50 "We had this mapping values are not allowed here" on YAML tag

I have a .yaml file open in Netbeans 8.1 which looks like the following: --- rules: - ltp aoi topology TopologyRule labels: - empty_A output: - entry_B Netbeans has a ...

netbeans

yaml

asked Jan 23 at 22:10

karobar

186 2 19

1 answer

102 views

11 votes

+50 Webpack - Style sheet loading but no fonts

I am able to see my stylesheet in the page without problems. However I cannot get my webfonts to work. I have tried to save the output of my css which doesn't happen. I believe this is why the fonts ...

javascript

css

fonts

sass

webpack

asked Jul 22 at 0:58

Jamie Hutber

8,671 17 57 120

2 answers

131 views

11 votes

+100 Android emulator failed to start after 360 seconds

I have Jenkins 1.568 installed on a Macbook Air running Ubuntu 14.04. I have the android emulator plugin installed, and the configuration I have set up runs the emulator in -no-window mode before ...

android

jenkins

asked Jun 27 '14 at 19:19

jwir3

2,615 3 21 53

4 answers

972 views

407 questions

Related Tags

android × 53

javascript × 43

ios × 37

python × 32

java × 31

c# × 31

jquery × 15

objective-c × 14

swift × 14

angularjs × 13

more related tags

Hot Network Questions

What does being "Linear" mean for a transformation and a function intuitively/graphically?

17

NANYANG TECHNOLOGICAL UNIVERSITY | SINGAPORE

Extracting Job Openings from the Web

The screenshot shows a Microsoft Internet Explorer browser window displaying the OPUS International website. The address bar shows http://www.foodscience.com/jobs_midwest.html. The website has a navigation menu on the left with links like 'Welcome', 'About OPUS', 'Executive Staff', 'Job Listings', 'Resume Form', 'Job Hunt Hints', 'Academic Links', 'Science Fair Help', 'Industry Assocs.', 'FAQs', 'Contact Us', 'Site Map', and 'e-mail'. The 'Job Listings' link is highlighted with a red circle. The main content area shows a job listing for 'Ice Cream Guru' at foodscience.com. The job details are extracted into a summary box on the right, which lists the following information:

- Source: www.foodscience.com/jobs_midwest.html
- JobTitle: Ice Cream Guru
- Employer: foodscience.com
- JobCategory: Travel/Hospitality
- JobFunction: Food Services
- JobLocation: Upper Midwest
- Contact Phone: 800-488-2611
- DateExtracted: January 8, 2001
- OtherCompanyJobs: foodscience.com-Job1

The job listing text on the website is as follows:

Ice Cream Guru

If you dream of cold, creamy chocolate or gooey, gooey cookie, there's a great opportunity for you to maintain and expand this major corporation's high-end ice cream base. Will be based in the Upper Midwest for about a year. And hey, California here I come! Requires a BS in Food Science or dairy, plus ice cream formulation experience. Will consider entry level with an BS and an internship. Contact Susan at susan@foodscience.com 1-800-488-2611

The hidden structure of language

- We're going beneath the surface...
 - Not just string processing
 - Not just keyword matching in a search engine
 - Search Google on “tennis racquet” and “tennis racquets” or “laptop” and “notebook” and the results are quite different ... though these days Google does lots of subtle stuff beyond keyword matching itself
- We want to recover and manipulate at least some aspects of language structure and meaning

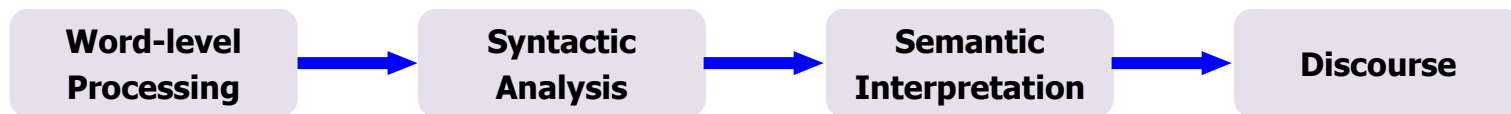


Major topics to cover

- Word-level processing:
 - Regular expression, Spelling correction, segmentation
 - Language models, POS tagging
- Syntax: knowledge of structural relation between words
 - Parsing
- Discourse: context

Applications:

- ◆ Information extraction, Named entity recognition
- ◆ Sentiment Analysis
- ◆ Machine translation



Example tasks (1)

- Word-level processing
 - Task 1: Locate all verbs and verbs only
 - E.g. the tower collapsed as a **result** of safety violations
 - *Is ‘result’ here a noun or a verb?*
- Syntactic processing
 - Task 2: Answer “Who killed John?”
 - E.g. “Mary killed John.”
 - E.g. “John was killed by Mary.”
 - E.g. “The guy who loved Mary killed John.”
 - E.g. “Mary is not sure of who killed John.”
 - *Hint: find subject of ‘killed’ whose object is ‘John’*

Example tasks (2)

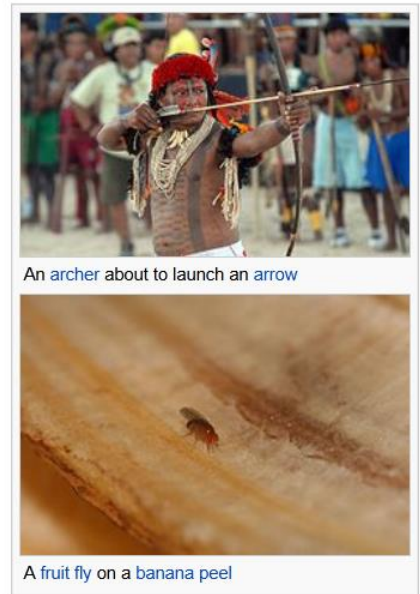
- Semantic processing
 - Task 3: Answer “Who killed John?”
 - E.g. Mary **assassinated** John.
 - Task 4: Answer “Who snores?”
 - E.g. **Everyone who** smokes snores and **John** smokes.
- Discourse analysis
 - Task 5: Answer “Who killed John?”
 - E.g. Mary threw **John** into sea. **He** drowned.

Learning Objective

- You will learn natural language processing at a **basic level**, establishing a solid understanding on the theory of **morphological**, **syntactic**, and semantic analysis.
- With that, you will gain skills to apply the NLP techniques to **real-world problems** by using **NLP packages** and **toolkits**.
- Upon completion of the course, you should be able to:
 - **Understand and analyze** the linguistic characteristics of written English
 - **Design and develop** a NLP system to analyze and process a general corpus
 - **Troubleshoot** for **domain-specific** NLP applications

Caveat

- Why NLP is difficult? NLP has an AI aspect to it.
 - The language is hugely **ambiguous**
 - We don't often come up with exact solutions/algorithms
- Example
 - Time *flies* like an arrow.
 - Fruit *flies* like a banana.
- What is “Java”?
 - [https://en.wikipedia.org/wiki/Java_\(disambiguation\)](https://en.wikipedia.org/wiki/Java_(disambiguation))



Ambiguity is Pervasive

- Find at least 5 meanings of this sentence: I made her duck
 - “**duck**” (lexical category): can be a noun or verb
 - “**her**” (lexical category): can be a possessive (“of her”) or dative (“for her”) pronoun
 - “**make**” (lexical semantics): can mean “create” or “cook”, and about 100 other things as well
- ✓ I cooked waterfowl for her
 - ✓ I cooked waterfowl belonging to her
 - ✓ I created the (plaster?) duck she owns
 - ✓ I caused her to quickly lower her head and body
 - ✓ I waved my magic wand and turned her into undifferentiated waterfowl

Language is still the ultimate UI (Example: Siri)



More about Assignment

- 4-5 members in a group
- Assignment release (before recess week)
 - Will have room for creativity
- Can use any programming language(s)
 - Can use any third-party libraries
- What you need to deliver:
 - Source code + final report (End of week-12)



Course Web Page

- The course web page can be found at **NTULearn**
- It will have the lecture notes, announcements, etc.
 - Slides cannot replace the textbook.
 - They are at most a guideline.

