,class 01

מבהואעלעיבודשפותעברית

יואבגולדברג,רועותארפתי,עדידדגן
What is NLP?

Algorithms that operate on Human Language
What is NLP?

- Takes as input text in human language and process it in a way that suggests an intelligent process was involved.
- Non-trivial useful output.
מרציו הקורסים / מיナンפים

פרופ עדי דגן

פרופ יואב גולדברג

פרופ רועת ארפחין
מרציו הקורות / מי אנחתו

פרופ יואב גולדברג
פרופ עודו דגן
פרופ נעית הרצטינ

הבנה עמוקה של מבני שפה + מודלי לימדה מוכנה
מרציו הקורסים / מי أنحاء

프וף יואב גולדברג

הננות NLP לא ממחיים

שיטות "למידה עמוקה":

תיאורית, פרשניית, הטיית

אבני בניין של הבנת שפה

תיעוד ובשפת טעויות

הכנה לומדה של מבני שפה

פרוייקטים-transparent של יישומי פעולות ביןorchards

 PROF. ruth ARTI

PROF. IOAV GOLDBERG

PROF. EIDU DGAN

מתודת אוטומטים

ווינטזרקטיב

חולם דעות משפטים

ייצוג דעות טקסטואלי

מבוסס שלאלת

עברית NLP

טכנית בשפת טעויות

הכנה לומדה של מבני שפה

פרוייקטים-transparent של יישומי פעולות ביןorchards

PROF. IOAV GOLDBERG

PROF. EIDU DGAN

PROF. RUTH ARTI
מי אתם?

מי שבתואר ראשון, בווס
raise hand
מי אtems?

מי שבתואר שני, בום raise hand בום
מי א홍?

שנライיט שומע השלמות/ת לשופות שאתם יודיעו שאינן עברית.
מי אתם?

שני בחורה
מי האם?  

מי שאהב سيוגרי לשו/שפה בתינוק: 

raise hand
מי אתס?

raise hand

מי صلى ליום ליום ולשון/שפה בשנייה?
מי אהבו?

למה בחירתם בקורס הזה? (כתבו בצ'אט)
קורונה

• קורס מקוון
•ﭑ벌 אנחננ ממליצים מאת להנייל להרצאות במעדו.

נגסה לשלב אינטראקציה בהרצאות
• אנו מצפים מכמ להשתתף ב

השתתפי בהרצאות
• תעוןسألת שלנה.
• תשאלسألת [לפי הפרוטוקול של המרצא גזר]

• מבקשים שתperate את המשלמה.
Introduction to Natural Language Processing

Semester A, 2020-2021

Instructors: Prof. Reut Tsarfaty, Prof. Yoav Goldberg and Prof. Ido Dagan
Emails: reut.tsarfaty@biu.ac.il; yogo@cs.biu.ac.il; dagan@cs.biu.ac.il
Offices: Building 216
Office Hours: By email appointment

This week's class

Zoom Link (NOTE: this may change between weeks.)

Course Requirements

Requirements

Resources

Online Discussion Forums

Managed via Piazza
Introduction to Natural Language Processing

Semester A, 2020-2021

Instructors: Prof. Reut Tsarfaty, Prof. Yoav Goldberg and Prof. Ido Dagan
Emails: reut.tsarfaty@biu.ac.il; yogo@cs.biu.ac.il; dagan@cs.biu.ac.il
Offices: Building 216
Office Hours: By email appointment

This week's class

Zoom Link (NOTE: this may change between weeks.)

Course Requirements

Requirements

Resources

Online Discussion Forums

Managed via Piazza
Introduction to Natural Language Processing

Semester A, 2020-2021

Instructors: Prof. Reut Tsarfaty, Prof. Yoav Goldberg and Prof. Ido Dagan
Emails: reut.tsarfaty@biu.ac.il; yogo@cs.biu.ac.il; dagan@cs.biu.ac.il
Offices: Building 216
Office Hours: By email appointment

This week's class

Zoom Link (NOTE: this may change between weeks.)

Course Requirements

Requirements

Resources

Online Discussion Forums

Managed via Piazza
Introduction to Natural Language Processing

Semester A, 2020-2021

Instructors: Prof. Reut Tsarfaty, Prof. Yoav Goldberg and Prof. Ido Dagan
Emails: reut.tsarfaty@biu.ac.il; yogo@cs.biu.ac.il; dagan@cs.biu.ac.il
Offices: Building 216
Office Hours: By email appointment

This week's class

Zoom Link (NOTE: this may change between weeks.)

Course Requirements

Requirements

Resources

Online Discussion Forums

Managed via Piazza

private message to "Instructors"
דרישות הקורס

• עבודה בית (40 אחוז מהציון)
  • 4 עבודות
• תכנון + כתיבת ד"ר"ח
• עבודה בזוגות
• דרישות חשוקה רצינית
• בחרה (60 אחוז מהציון)
דרישות הקורס

הקורס

דרישות

• העבדות ביצת (40 אחות מהזמנון)

• עבדות 4 עבדות

• תכונות + כתיבת ד”ר

• עבדות בוגרות

• דורשיםanken השקפה רצינית

• בהינתן (60 אחות מהזמנון)

הкова הליגה占领 העבדות ואת כל העבדות כדי לעבור את הקורס.

הкова לעבור את המבחן כדי לעבור את הקורס.

אסור לעתיך. (גם בעבדות עבדות בבותינה. לא מחברים, לא משנים קודמות. לא מחברים. גם לא חלקי קוד او חלקי תשובה.)
דרישות הקורס

עדכון:

• יש להגיש על פי הھנחיות. יְש לוֹדוֹא שֶהָכֹדּ מַחְקִּמֵּל וּכְרֵי.

• תַּרְגִּיל שֶלָּא יְגֵשַׁלַּא הַהִנָּחִיָּה, לָא יְבָדָק / יְכָבֵל יִיוָנָן נַבְלָל.

• יש להגיש את העבדות חָמָן.

• לכְּלָ שְׁטֹוֹדְנֵנָי תִּשׁ 10 "נַיֵּאִיחָוּ" שֶאֶפָּשַׁר לַפְּרוֹשׁ עֵלַ פַּנֶּי כָּל

עַבְדֻדוֹת. פְּרֵיִם בַּאֲחָר הַקורֵיס הַחָת.
Introduction to Natural Language Processing

Semester A, 2020-2021

Instructors: Prof. Reut Tsarfaty, Prof. Yoav Goldberg and Prof. Ido Dagan
Emails: reut.tsarfaty@biu.ac.il; yogo@cs.biu.ac.il; dagan@cs.biu.ac.il
Offices: Building 216
Office Hours: By email appointment

This week's class

Zoom Link (NOTE: this may change between weeks.)

Course Requirements

Requirements

Resources

Online Discussion Forums

Managed via Piazza
דרישות הקורס

מחלק הוראות:

• בוחנים:
  • שאלות על החומר שנלמד, כדי לעזר ול海淀, ולראות מה הבננת.

• לא ישפיעו על הצינו.
  • מבㅅים מכילים להשתתף.
턴שורות עם צוות הקורס

• אנא לא לפנות לעוזר הקורס באימייל, אנו נגיני לא גнии.

Piazza - כל התנשואות עם צוות הקורס יתבצעו במועצה ה-

• ההדעת שלרגיטיות לכלול: נא להזדהות כ unheard פומבי.

• elaاهرם בקורס: נא להזדהות, אם רלוונטי.

• ההדעת אישיות: שלולות ל- Instructors- בפרטי, כר שלכל צוות הקורס יהייה מכתובה.
חומרי קריאה

• ידיעת מאמרים או תכניות הרחבת באחת.
• יש קישור לשפרים момלצים באחת.
קורסים נוספים

• שיטות דימ לרגלי ולקטטים במחפסים ורפרפים (סמטער ב', גולדברג)

• سمינריוויי "ממבטיה למשמעות" (סמטער א', אהרוני)

• سمינריוויי מתכדמים בחבון.metro (סמטער ב', דגונ)

• سمינר ישומי למידת מכונה בעיון של שפה ממידה (סמטער ב', דגונ)

• سمינריוויי באלאגוריתמי למידה מיישב שפה (סמטער ב', גולדברג)

• תרגום מכונה (סמטער ב', אהרוני)
עד כאן רקע.

•(QLORS
[מעבר לעידה]
שפת תביעה זה קשה

אם זה מה שמהÅת הגדול:
• עיבוד שפה זה קשה.
• למה זה קשה?
  (ולמה זה מעניין)
• איך, בגדול, מתמודדים עמו זה.
החייםマルアイマルיאס פשׁרוּת

הדבר השני שטיקו המקורס:

הכל הוה פשה

מה הוא אומר "לותברميلים למשפטיים"?

לתבר שלאלות لتحXObject עבש "מלק הכרויה" (וטסו)

לענהן תשובות לשירות לקוחות?

שיהיה תופשיית בנסאה מות לאויר?

שיהיה תופשיית על הכל (מבווח טוירינגו)?

אולי רכ משלטימי עד אורי 10 מלילים?

אולי פסקאות?
ＮＬＰ
בנירת מערכה

• לברוח משורת סבירות
• להבינה שישינו פשרה
• להבינה מה הפשרה שישינו
• لنסח לשפר, אם צרי, עד כמות שאריה.
the basic units of text processing
What are we processing?

Document Collection / Cluster

- Document
- Section
- Paragraph
- Sentence
- Word
What are we processing?

Document Collection / Cluster

- Document
- Section
- Paragraph
- Sentence
- Word

by date
by topic
by web domain
...

BI UNL P
What are we processing?

Document Collection / Cluster

Document

Section

Paragraph

Sentence

Word
What are we processing?

Document Collection / Cluster

- Document
- Section
- Subsection
- Paragraph
- Sentence
- Phrases...
- Word
What are we processing?

Document Collection / Cluster
Document
Section
Paragraph
Sentence
Subsection
Phrases...
Word
Characters, morphemes
What are we processing?

Document Collection / Cluster

- Document
  - Section
  - Paragraph
  - Sentence
  - Word
What is a document?
British lawmakers clash ahead of Brexit vote 05:57

Story highlights

Britons will vote Thursday on whether their country should remain in the European Union

Henry Wismayer: Urge to leave has caught the wind of native disaffection

Editor's Note: Henry Wismayer is a writer and commentator based in London. You can follow him @henrywismayer. The views expressed are his own.

(CNN) — For any external observer curious to understand the issues shaping the debate over Britain's membership of the European Union, last Wednesday's edition of The Sun, the country's most popular daily newspaper, produced a revealing metaphor.

A day after the Rupert Murdoch-owned tabloid declared its support for the "Leave" campaign, its front-page went full-hysteria. A "biblical swarm" of super-moths has invaded Britain from mainland Europe, it exclaimed without a trace of irony, "and is set to annihilate our cabbages."

The prominence given to these ghastly foreign moths, accompanied on Page 2 with a helpful diagram depicting a battalion of Lepidoptera hovering over the English Channel, tells you all you need to know about Britain's EU referendum debate, where fear and frustration have come to overshadow the biggest political decision in a generation.
What is a document?

Story highlights

Britons will vote Thursday on whether their country should remain in the European Union

Henry Wismayer: Urg to leave has caught the wind of native disaffection

Editor's Note: Henry Wismayer is a writer and commentator based in London. You can follow him @henrywismayer. The views expressed are his own.

(CNN) — For any external observer curious to understand the issues shaping the debate over Britain’s membership of the European Union, last Wednesday’s edition of The Sun, the country’s most popular daily newspaper, produced a revealing metaphor.

A day after the Rupert Murdoch-owned tabloid declared its support for the “Leave” campaign, its front-page went full-hystera. A “biblical swarm” of super-moths has invaded Britain from mainland Europe, it exclaimed without a trace of irony, “and is set to annihilate our cabbages.”

The prominence given to these ghastly foreign moths, accompanied on Page 2 with a helpful diagram depicting a battalion of Lepidoptera hovering over the English Channel, tells you all you need to know about Britain’s EU referendum debate, where fear and frustration have come to overshadow the biggest political decision in a generation.
What is a document?

and maybe each paragraph is a document?

Britons will vote Thursday on whether their country should remain in the European Union.

Henry Wismayer: Urg to leave has caught the wind of native disaffection

(CNN) — For any external observer curious to understand the issues shaping the debate over Britain’s membership of the European Union, last Wednesday's edition of The Sun, the country's most popular daily newspaper, produced a revealing metaphor.

A day after the Rupert Murdoch-owned tabloid declared its support for the "Leave" campaign, its front-page went full-hysteria. A "biblical swarm" of super-moths has invaded Britain from mainland Europe, it exclaimed without a trace of irony, "and is set to annihilate our cabbages."

The prominence given to these ghastly foreign moths, accompanied on Page 2 with a helpful diagram depicting a battalion of Lepidoptera hovering over the English Channel, tells you all you need to know about Britain’s EU referendum debate, where fear and frustration have come to overshadow the biggest political decision in a generation.
What are we processing?

Document Collection / Cluster

Document

Section

Paragraph

Sentence

Word
What is a sentence?

sentences = text.split(".")?
What is a sentence?

(we'll get back to this)
What are we processing?

Document Collection / Cluster

Document
Section
Paragraph
Sentence
Word
What is a word?
What is a word?

sequence of characters

basic unit of meaning?

white-space tokenized?
What is a word?

doesn't

John's

unlucky
What is a word?

שלום
What is a word?

שלום
שלום
What is a word?

שלום
shalom
שלום
shal
אני
ayin
What is a word?
What is a word?

 syrup

they told  count!  his book  (ספר שלوها)
What is a word?

סיפרו

they told count! his book

(ספר של יואם)

he told it he₁ cut his₂ hair
What is a word?

Words aren't just defined by blanks.

Problem 1: Compounding

"ice cream", "website", "web site", "New York-based"

Problem 2: Other writing systems have no blanks

Chinese:

I start(ed) writing novel(s)

Problem 3: Clitics

English: "doesn't", "I'm"

Italian: "dirglielo" = dir + gli(e) + lo = tell + him + it

by Julia Hockenmaier
whitespace is not enough
What is a word?

- ice cream
- web site
- New York
- New York-Based
What is a word?

gave up

made sense

took a picture

took apart

took the toy apart
Case Study: Sentence Boundary Detection.
What is a sentence?

sentences = text.split("."),
A graphic adventure game is a form of adventure game. They are distinct from text adventures. Whereas a player must actively observe using commands such as "look" in a text-based adventure, graphic adventures revolutionized gameplay by making use of visual human perception. Eventually, the text parser interface associated with older interactive fiction games was phased out in favor of a point-and-click interface, i.e., a game where the player interacts with the game environment and objects using an on-screen cursor. In many of these games, the mouse pointer is context sensitive in that it applies different actions to different objects.
What is a sentence?

sentences = text.split(". ")?
Georg "Mr. George" Svendsen (19 March 1894 – 1966) was a Norwegian journalist and crime novelist. He was born in Eidanger, and started his journalistic career in Bratsberg-Demokraten before moving on to Demokraten where he was a subeditor. In 1921 he was hired in Fremtiden and replaced in Demokraten by Evald O. Solbakken. In 1931 he was hired in Arbeiderbladet. Under the pen name "Mr. George" he became known for his humorous articles in the newspaper. At his death he was also called "the last of the three great criminal and police reporters in Oslo", together with Fridtjof Knutsen and Axel Kielland. He was also known for practising journalism as a trade in itself, and not as a part of a political career. He retired in 1964, and died in 1966.

He released the criminal novels Mannen med ljåen (1942), Ridderne av øksen (1945) and Den hvite streken (1946), and translated the book S.S. Murder by Quentin Patrick as Mord ombord in 1945. He released several historical books: Rørleggernes Fagforenings historie gjennem 50 år (1934), Telefonmontørenes Forening Oslo, gjennem 50 år (1939), Norsk nærings- og nydelsesmiddelarbeiderforbund: 25-års beretning (1948), De tause vitner: av rettskjemiker Ch. Bruffs memoarer (1949, with Fridtjof Knudsen) and Elektriske montørens fagforening gjennom 50 år (1949).
Sentence Boundary Detection

- How would you solve this?

(discuss)
Sentence Boundary Detection

• Worth reading:

Sentence Boundary Detection and the Problem with the U.S.

Dan Gillick
Computer Science Division
University of California, Berkeley
dgillick@cs.berkeley.edu
What is a sentence?

It was a bad time. (it is always a bad time.)
What is a sentence?

In a quiet voice, he said "this will not work. I am quitting", and he left the room.
How often do you find yourself on the internet looking at the same boring pages? You know there is something out there but you don't know where to look.

Trust me, how bad could it be?

Click above for a new random website
Links will open in a new window

**Disclaimer:** The websites you are forwarded to may contain naughty stuff. There is no known "porn" but sites change hands often and sometimes they change ethical direction. If you see something that looks like porn or is otherwise unacceptable let us know and we'll decide if it should be removed or if you're just a crank who likes to complain.

Don't leave the house with your fly open. Make sure you've got up-to-date virus protection.

Colophon & Warning & Contact

Want to add your own links? After reading the **warning** type the url in the box below

[Box for URL input]

If you see something that shouldn't be here let us know

random website dot com 2001-2011

"I'd like to quit thinking of the present, like right now, as some minor insignificant preamble to something else."
How often do you find yourself on the internet looking at the same boring pages? You know there is something out there but you don't know where to look

Trust me, how bad could it be?

randomlink

Click above for a new random website
Links will open in a new window

Disclaimer: The websites you are forwarded to may contain naughty stuff. There is no known "porn" but sites change hands often and sometimes they change ethical direction. If you see something that looks like porn or is otherwise unacceptable let us know and we'll decide if it should be removed or if you're just a crank who likes to complain.

Don't leave the house with your fly open. Make sure you've got up-to-date virus protection.

Colophon & Warning & Contact

Want to add your own links? After reading the warning type the url in the box below

If you see something that shouldn't be here let us know

random website dot com 2001-2011

"I'd like to quit thinking of the present, like right now, as some minor insignificant preamble to something else."
How often do you find yourself on the internet looking at the same boring pages? You know there is something out there but you don’t know where to look.

Trust me, how bad could it be?

Click above for a new random website
Links will open in a new window

Disclaimer: The websites you are forwarded to may contain naughty stuff. There is no known "porn" but sites change hands often and sometimes they change ethical direction. If you see something that looks like porn or is otherwise unacceptable let us know and we’ll decide if it should be removed or if you’re just a crank who likes to complain.

Don’t leave the house with your fly open. Make sure you’ve got up-to-date virus protection.

Colophon & Warning & Contact

Want to add your own links? After reading the warning type the url in the box below

If you see something that shouldn’t be here let us know

random website dot com 2001-2011

"I’d like to quit thinking of the present, like right now, as some minor insignificant preamble to something else."
How often do you find yourself on the internet looking at the same boring pages? You know there is something out there but you don't know where to look.

Trust me, how bad could it be?

Randomlink

Click above for a new random website. Links will open in a new window.

Disclaimer: The websites you are forwarded to may contain naughty stuff. There is no known "porn" but sites change hands often and sometimes they change ethical direction. If you see something that looks like porn or is otherwise unacceptable let us know and we'll decide if it should be removed or if you're just a crank who likes to complain.

Don't leave the house with your fly open. Make sure you've got up-to-date virus protection.

Colophon & Warning & Contact

Want to add your own links? After reading the warning type the url in the box below.

If you see something that shouldn't be here let us know.

Random website dot com 2001-2011

"I'd like to quit thinking of the present, like right now, as some minor insignificant preamble to something else."
Sentence Boundary Detection

- Perhaps the most basic task.
- Non-trivial.
- Need to consider corpus, features, annotation procedure, biases.
- Need to think of your use-cases, choose your sentence definition, methods, and trade-offs.