



TEXT MINING & NATURAL LANGUAGE PROCESSING

Buu NGUYEN

INTRODUCTION

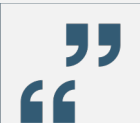
- In this study, we analyzed the data of 50 doctoral dissertations and 7000 phrases of English.
- The goal was to successfully apply the basics of text mining and natural language processing techniques such as Bag-of-Words and Part-of-Speech Tagging.



METHODS



The data was handled by the instructors.



Python and its libraries such as **nlTK**, **sklearn**, and **spacy-stanza** was used.

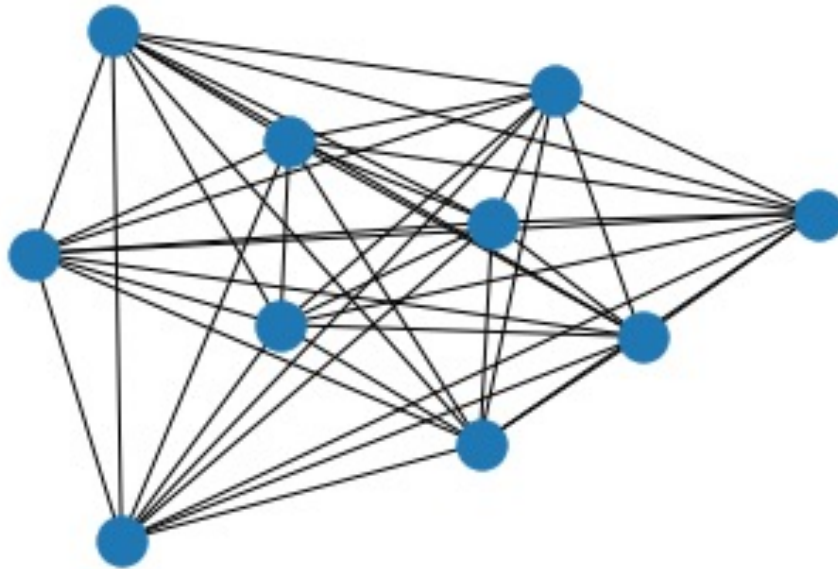
METHODS

The text data from 50 doctoral dissertations were read as pre-processing. Then the cosine similarity between dissertations was calculated and plotted into a graph.

7000 English phrases were extracted from the dataset. The POS-Tagging technique was used to detect the verbs inside each phrase. Then, a multi-choice question database was created for further studies.

RESULTS

Cosine similarity network



Each nodes is a doctoral dissertation.

Distances between nodes are the cosine similarities between them.

Phrase	Question	Lemma	Answer	D1	D2	D3	F_Answer	F_D1	F_D2	F_D3
The beauty of the landscape struck the travellers.	The beauty of the landscape ... the travellers.	strike	struck	strikes	striking	strided	past tense	3rd person singular present	present participle	past tense
Nobody knows the truth about this affair.	Nobody ... the truth about this affair.	know	knows	knowed	knowing	knows	3rd person singular present	past tense	present participle	3rd person singular present
He did not help you out of kindness.	He did not ... you out of kindness.	help	help	helps	help	helping	infinitive	3rd person singular present	infinitive	present participle
His wickedness had no limits.	His wickedness ... no limits.	have	had	have	haves	haved	past tense	infinitive	3rd person singular present	past participle
His elegance impressed the assembly.	His elegance ... the assembly.	impress	impressed	impresses	impress	impressed	past tense	3rd person singular present	infinitive	past participle
There is a big difference between the western and the asiatic culture.	There ... a big difference between the western and the asiatic culture.	be	is	bing	bes	bed	3rd person singular present	present participle	3rd person singular present	past participle
He has high ideals.	He ... high ideals.	have	has	haved	have	haved	3rd person singular present	past participle	infinitive	past tense
He was struck by the modernity of the underground.	He ... by the modernity of the underground.	strike	was struck	strided	strike	strided	past participle	past tense	infinitive	past participle

RESULTS

QUESTION 1
MULTIPLE CHOICE

Fill in
the blank

The cat is too high,
I cannot _____ it.

- ☐ A Reach
- ☐ B Reaches
- ☐ C Reached
- ☐ D Reaching

QUESTION 1

ANSWER

Fill in
the blank

The cat is too high,
I cannot reach it.



- ☒ A Reach
- ☐ B Reaches
- ☐ C Reached
- ☐ D Reaching