

Text Mining

Jing-Doo Wang, Ph.D.

Associate Professor

CSIE Department of Asia University, Taiwan

jdwang@asia.edu.tw

Time: Tue, 9:10am~12:00am, Room: 1627
Room:1517, ext:1847



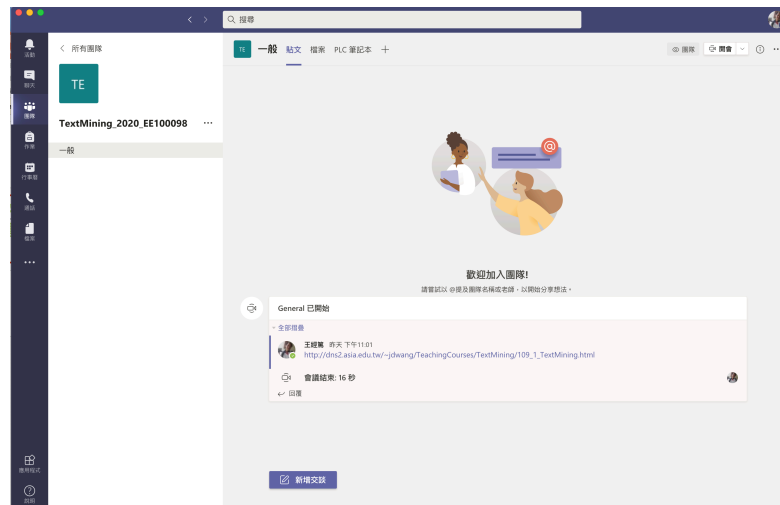
http://dns2.asia.edu.tw/~jdwang/TeachingCourses/TextMining/109_1_TextMining.html

Pre-Requests

- ▶ Python Programming or any other programming languages
- ▶ [Learn and Master Python in a Month](#) (From: [Jhankar Mahbub](#), 2019)
- ▶ [Online Python Course for Beginners](#) (edx)

TextMining_2020_EE100098 (Virtual Exchange Program)

- ▶ You should use the student ID of Asia university to login (MS_Team).
- ▶ https://teams.microsoft.com/dl/launcher/launcher.html?url=%2F_%23%2F%2Fteam%2F19%3A2a839bbb700043debf0e420a16d15e81%40thread.tacv2%2Fconversations%3FgroupId%3D8003837e-3617-4d54-bb06-6fe69b2ed6cd%26tenantId%3Dda140a55-a3ec-49de-889e-93c4e142a2fb&type=team&deeplinkId=fa8e8252-f8f1-4973-8a56-d9156908df25&directDl=true&msLaunch=true&enableMobilePage=true&suppressPrompt=true



Web Page for this course

http://dns2.asia.edu.tw/~jdwang/TeachingCourses/TextMining/109_1_TextMining.html

Text Mining

[TextMining_2020_EE100098 \(MS Team\)](#)

Time: Tue, 9:10am~12:00am, Room: I627
jdwang@asia.edu.tw, Room:I517, ext:1847

Pre-Requests

Some knowledge of Python Programming or any other programming languages

[Score](#)

Text Books

[Text Analytics with Python: A Practitioner's Guide to Natural Language Processing \(2nd version\)](#) Sarkar, Dipanjan,2019-05-22,ISBN-13:9781484243534
[code-downloads](#)
[Book_TextAnalyticswithPython_Notes.html](#)

References books

- [Hands-On Python Natural Language Processing](#) Aman Kedia , Mayank Rasu, 2020-06-26, ISBN-13:9781838989590
[code-downloads](#)
[code-downloads \(github\)](#)
- [Python Data Science Handbook \(2015\)](#) Jake VanderPlas
- [Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow, 2nd Edition\(2017\)](#) Sebastian Raschka, Vahid Mirjalil

Text Book

- ▶ [Text Analytics with Python: A Practitioner's Guide to Natural Language Processing \(2nd version\)](#)
- ▶ Sarkar, Dipanjan, 2019-05-22, ISBN-13: 9781484243534
- ▶ [code-downloads](#)
- ▶ [Book_TextAnalyticswithPython_Notes.html](#)



References books

- ▶ Hands-On Python Natural Language Processing Aman Kedia , Mayank Rasu, 2020-06-26, ISBN-13:9781838989590
code-downloads
code-downloads (github)
- ▶ Python Data Science Handbook (2015) Jake VanderPlas
- ▶ Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow, 2nd Edition(2017) Sebastian Raschka, Vahid Mirjalil

Weeks Date Course content (18 weeks)

- ▶ 1 2020/9/22 Introduction
- 2 2020/9/29 Python for NLP
- 3 2020/10/6 Text Preprocessing
- 4 2020/10/13 Traditional Feature Engineering Models
- 5 2020/10/20 Advanced Feature Engineering Model
- 6 2020/10/27 Text Classification-Vectorization
- 7 2020/11/3 Text Classification-Classifiers
- 8 2020/11/10 Middle Project Presentation
- 9 2020/11/17 Middle Project Report
- ▶ 10 2020/11/24 Text Similarity
- 11 2020/12/1 Text Clustering
- 12 2020/12/8 Text Cluster- Visualization
- 13 2020/12/15 Text Mining vs. Deep Learning(1)
- 14 2020/12/22 Text Mining vs. Deep Learning(2)
- 15 2020/12/29 AWS Educate program and AWS Academy program
- 16 2021/1/5 Text Mining vs. AWS
- 17 2021/1/12 Final Project presentation: Text Mining + AWS
- 18 2021/1/19 Final Project (Report): Text Mining + AWS

Grade

<https://drive.google.com/file/d/1SinQRWeFR63NH4hKuX7-FtWE3uZX6Orl/view>

- ▶ Attendance (10%):
 - ▶ 10% if you attend every week on time, Absent (-1/per week), Late (-0.5/per week),
 - ▶ at most -10.
- ▶ Homework 1 (15%=> 10%):
 - ▶ (2020/10/27 report (word(pdf)+YouTube (URL: sharing)) to Moodle)
- ▶ Middle Project (30%):
 - ▶ 2020/11/10, presentation (ppt)
 - ▶ 2020/11/17 report (word(pdf)+YouTube (URL: sharing)) to Moodle)
- ▶ Homework 2 (15%=> 10%): :
 - ▶ (2020/11/22 report (word(pdf)+YouTube (URL: sharing)) to Moodle)
- ▶ Final Project (40%):
 - ▶ 2021/1/12, presentation (ppt),
 - ▶ 2021/1/19 report (word(pdf)+YouTube (URL: sharing)) to Moodle)

A	B	C	D	E	F	G	H	I	J
				10%	10%	30%	10%	40%	100%
學生系所	學號(ID)	姓名(Name)	Country	Attendance: Absent (-1/per week), Late (-0.5/per week,)	HW1	Middle Project	HW2	Final Project	Score
資訊系	108221005	Agung Mulyo Widodo	IN	10					10
		Afin Pradana	Indonesia	10					10
		Bastian Farandy	Indonesia	10					10
		Bobbi Aditya	Indonesia	10					10
		Iffa Ihan Nabila	Indonesia	10					10
		Muhammad Fath Fathulqishan Wisnu	Indonesia	10					10
		M Thakut Nadhif G.	Indonesia	10					10
		Jeremy Vigny Wongso	Indonesia	10					10
		Reza Adiputra Maranatha	Indonesia	10					10
		Kholilatul Amalah	Indonesia	10					10
		VISHNU VARDHAN CHANDRAGIRI	India	10					10
		Ravi Teja Gorantla	India	10					10
		Ethamakkala Granendra Naidu	India	10					10
		Thales' Goyatiri	India	10					10
		Brijmaddy Bhavitha	India	10					10
		Baljan Shuk	India	10					10
		SAPARIM GNANA SRINIDI	India	10					10
		MARHALA PRAVEEN KUMAR REDDY	India	10					10
		V. Madhana Teja	India	10					10
		HIMA VIGNESH MALICHETTI	India	10					10
		BILLU BHARGAV	India	10					10
		P ARUN TEJA	India	10					10
		Bkaushalya	India	10					10

References

http://dns2.asia.edu.tw/~jdwang/TeachingCourses/TextMining/TExtMining_References.html

- ▶ [NLTK Corpora](#)
- ▶ [Learn and Master Python in a Month](#)
- ▶ [Python 從入門到精通：一個月就夠了！](#)
- ▶ [Text Mining in Python: Steps and Examples](#)(Dhilip Subramanian, Aug 22, 2019)
- ▶ [\(Tutorial\) Text ANALYTICS for Beginners using NLTK](#)(Avinash Navlani, December 14th, 2019)
- ▶ [Text Mining & 網路爬蟲 web crawler | Google新聞與文章文字雲 | Python](#)(2019-03-03 JAMLEECUTE)
- ▶ What are the differences between [Occidental Languages](#) and [Oriental Languages?](#)
[Applied Natural Language Processing | Using AI to Build Real Products](#)(Arushi Raghuvanshi)

Computational Platform in AWS

- ▶ To someone without sufficient computing power when handling with Big Data
- ▶ [AWS Educate Program \(Portal Site\)](#)
- ▶ [Apply for an AWS Educate \(By Suca\)](#)
- ▶ [AWS Educate Program Information](#)

Thanks for your attentions