SYD

(Sort Your Docx)

2201792354 - Edward Susanto

2201730293 - Nicholas Evan Prawiro

2201765353 - Vicky Joe

Library:

Algorithmia

Python-docx

How:

1. Insert Your document into a folder and copy the folder path ex:

C:\Users\Asus\Documents\BINUS\NLP

2. Paste the folder Path (YOUR PATH) inside LDetection.py

like this

Ex:

organize\_files\_by\_language(r'C:\Users\Asus\Documents\BINUS\NLP\ProjectNLP')

3. Run the .py file

4. open the folder location

5. Tadaaa your docx is sorted

Some of The supported languages are:

ast Asturian

bn Bengali

da Danish

de German

el Greek

en English

es Spanish

et Estonian

eu Basque

fa Persian

fi Finnish

fr French

id Indonesian

is Icelandic

it Italian

ja Japanese

ko Korean

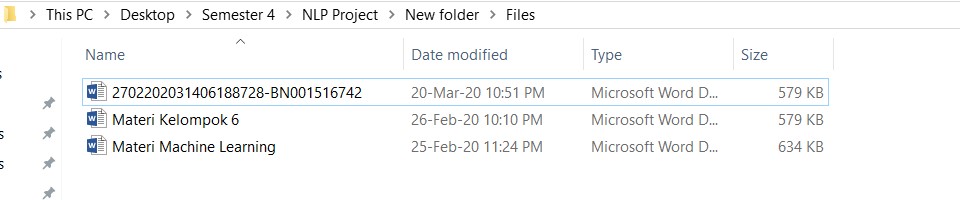
ms Malay

zh-cn Simplified Chinese

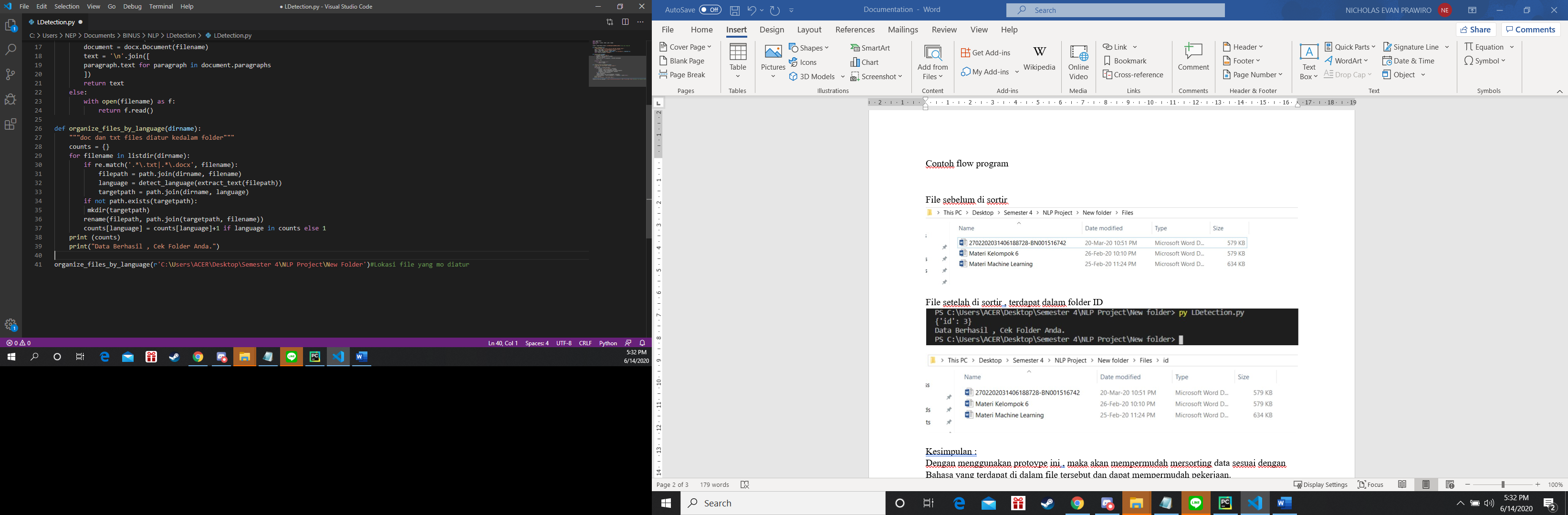
Penjelasan singkat

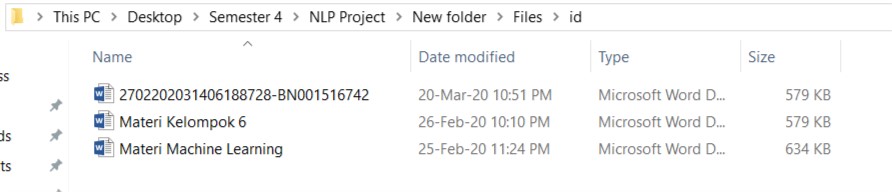
Code yang dibuat menggunakan API dari algorithmia yang dimana code tersebut langsung menggunakan fungsi-fungsi yang disediakan oleh algorithmia sehingga lebih mudah, yang dimana pada file-file yang ada di extract isinya kedalam sebuah variable text kemudian di process untuk melihat bahasa apa yang digunakan, yang akan di kembalikan berupa code ISO 639, setelah itu akan di masukan kedalam satu folder sesuai dengan bahasanya.

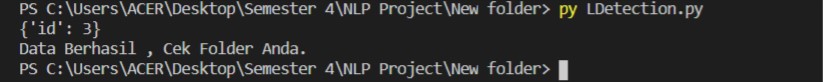
Contoh flow program

File sebelum di sortir

Kemudian ubah organize\_file\_by\_language



File setelah di sortir , terdapat dalam folder ID



Kesimpulan :

Dengan menggunakan protoype ini , maka akan mempermudah mersorting data sesuai dengan Bahasa yang terdapat di dalam file tersebut dan dapat mempermudah pekerjaan.

Test data Source : <https://www.duxburysystems.com/documentation/dbt12.4/samples/word/samples.htm>

Link Video :

<https://www.youtube.com/watch?v=-gOwDPklhWA>

DAFTAR PUSTAKA

algorthmia.com .(2019,October 03).Python

.<https://algorithmia.com/developers/clients/python>

algorithmia.com.(2019, June 12).Language Identification.

<https://algorithmia.com/algorithms/nlp/LanguageIdentification>

python-docx.com(2013).Python Docx Documentation.

<https://python-docx.readthedocs.io/en/latest/user/documents.html>

docs.python.org(2013, June 12).Miscellaneous operating system interfaces.

<https://docs.python.org/3/library/os.html>

docs.python.org(2013, June 12).Miscellaneous operating system interfaces.

<https://docs.python.org/3/library/re.html>

Sample MS-Word Files in Many languages. (n.d.). Retrieved June 16, 2020, from <https://www.duxburysystems.com/documentation/dbt12.4/samples/word/samples.htm>