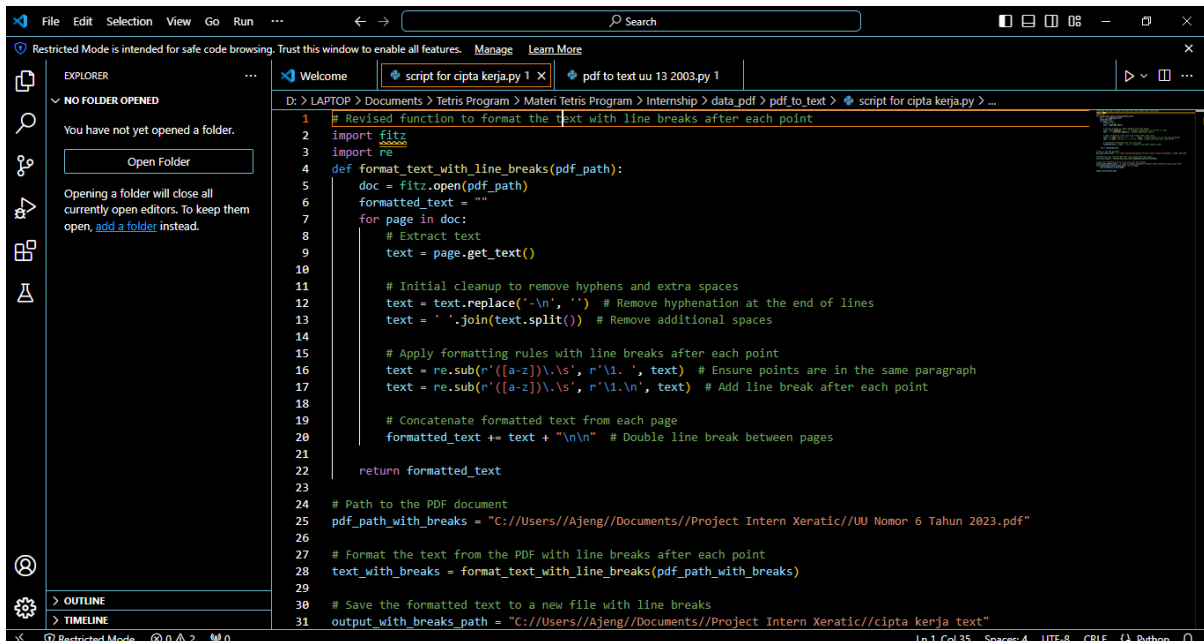


USER GUIDE

DATA PREPARATION SCRIPT: UU KETENAGAKERJAAN

A. Penggunaan Formatting PDF to Text Script

1. Buka File *script for cipta kerja.py* untuk ubah file *UU Nomor 6 Tahun 2023.pdf* atau file *pdf to text uu 13 2003.py* untuk ubah file *UU Nomor 13 Tahun 2003.pdf* di Visual Studio Code



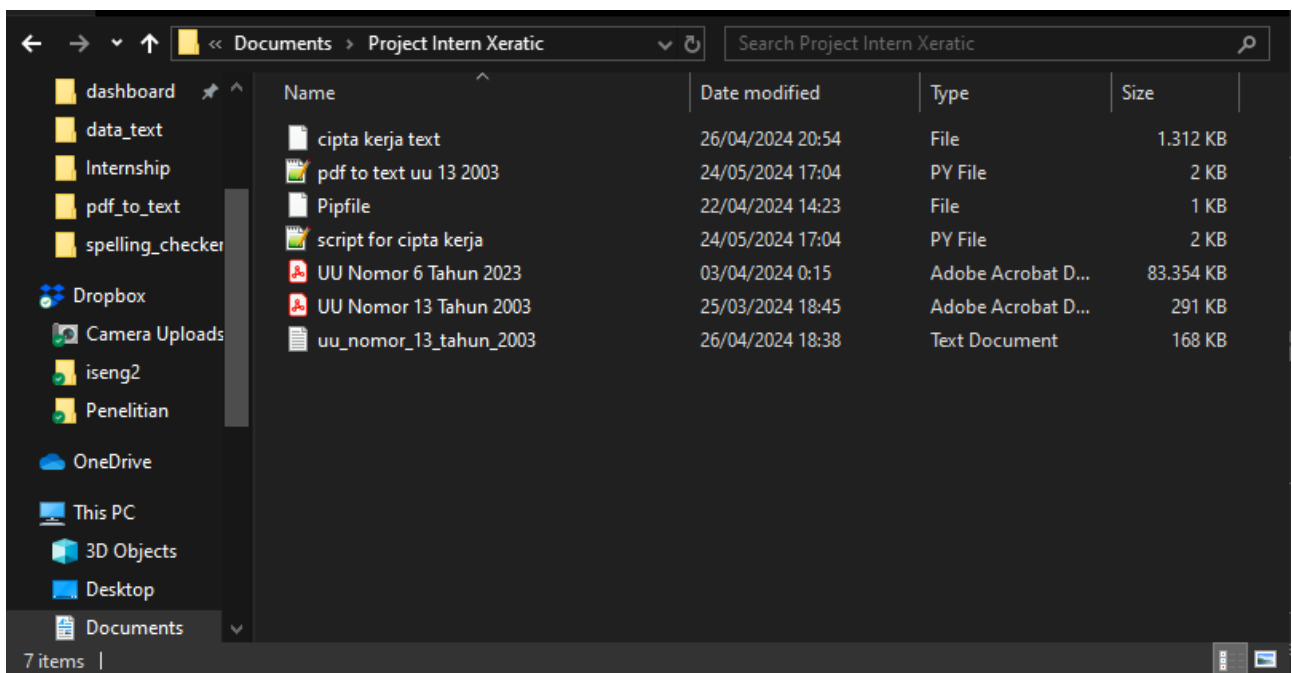
The screenshot shows the Visual Studio Code interface with a Python script open. The script is titled 'script for cipta kerja.py' and is located at 'D:\LAPTOP > Documents > Tetris Program > Materi Tetris Program > Internship > data_pdf > pdf_to_text > script for cipta kerja.py'. The script defines a function 'format_text_with_line_breaks(pdf_path)' that uses the 'fitz' library to open a PDF, extract text from each page, and format it with line breaks. The script then uses this function to format a specific PDF file and saves the result to a new file.

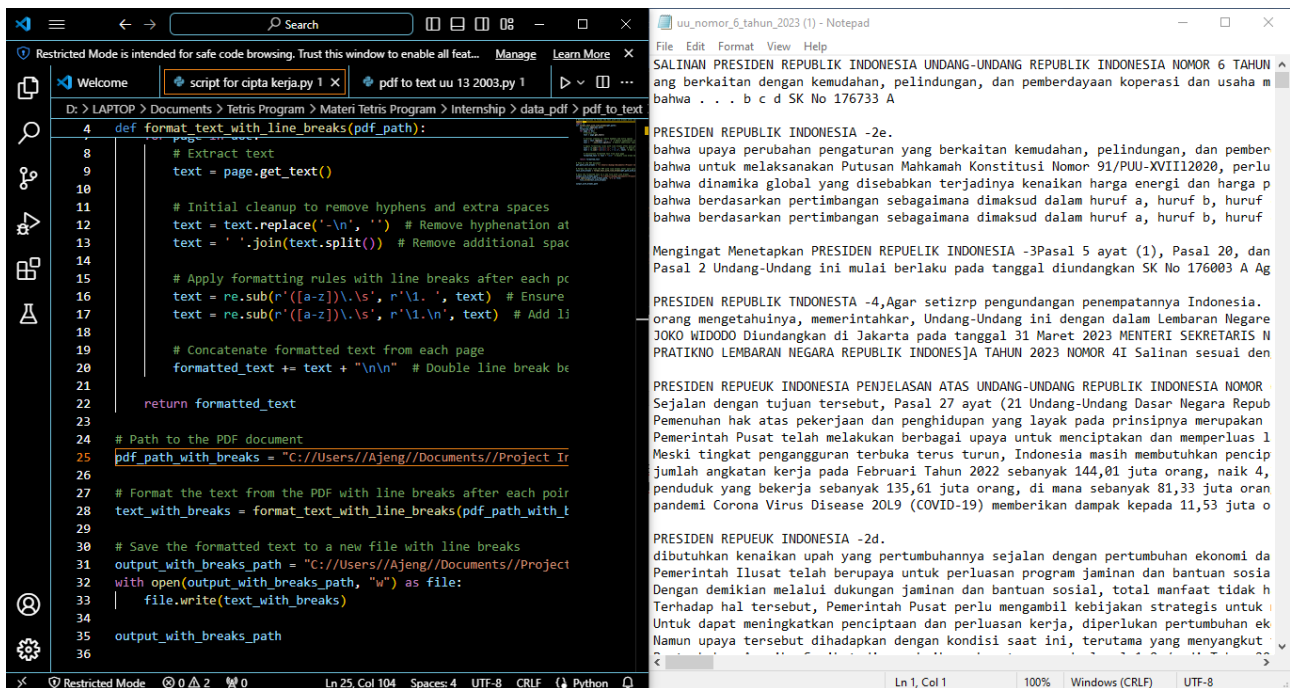
```
1 # Revised function to format the text with line breaks after each point
2 import fitz
3 import re
4 def format_text_with_line_breaks(pdf_path):
5     doc = fitz.open(pdf_path)
6     formatted_text = ""
7     for page in doc:
8         # Extract text
9         text = page.get_text()
10
11         # Initial cleanup to remove hyphens and extra spaces
12         text = text.replace('\n', '') # Remove hyphenation at the end of lines
13         text = ' '.join(text.split()) # Remove additional spaces
14
15         # Apply formatting rules with line breaks after each point
16         text = re.sub(r'([a-z])\.s', r'\1. ', text) # Ensure points are in the same paragraph
17         text = re.sub(r'([a-z])\.s', r'\1.\n', text) # Add line break after each point
18
19         # Concatenate formatted text from each page
20         formatted_text += text + "\n\n" # Double line break between pages
21
22     return formatted_text
23
24 # Path to the PDF document
25 pdf_path_with_breaks = "C://Users//Ajeng//Documents//Project Intern Xeratic//UU Nomor 6 Tahun 2023.pdf"
26
27 # Format the text from the PDF with line breaks after each point
28 text_with_breaks = format_text_with_line_breaks(pdf_path_with_breaks)
29
30 # Save the formatted text to a new file with line breaks
31 output_with_breaks_path = "C://Users//Ajeng//Documents//Project Intern Xeratic//cipt kerja text"
```

2. Ubah path sesuai folder dimana file PDF yang akan diubah ke file text berada pada script (line 25):
pdf_path_with_breaks = "C://Users//Ajeng//Documents//Project Intern Xeratic//UU Nomor 6 Tahun 2023.pdf"
3. Ubah path sesuai folder dimana file text yang telah diubah akan ditempatkan pada script (line 31):
output_with_breaks_path = "C://Users//Ajeng//Documents//Project Intern Xeratic//cipt kerja text"

```
4 def format_text_with_line_breaks(pdf_path):
5     # Extract text
6     text = page.get_text()
7
8     # Initial cleanup to remove hyphens and extra spaces
9     text = text.replace('-', '') # Remove hyphenation at the end of lines
10    text = ' '.join(text.split()) # Remove additional spaces
11
12    # Apply formatting rules with line breaks after each point
13    text = re.sub(r'([a-z])\.\s', r'\1. ', text) # Ensure points are in the same paragraph
14    text = re.sub(r'([a-z])\.\s', r'\1.\n', text) # Add line break after each point
15
16    # Concatenate formatted text from each page
17    formatted_text += text + "\n\n" # Double line break between pages
18
19    return formatted_text
20
21 # Path to the PDF document
22 pdf_path_with_breaks = "C://Users//Ajeng//Documents//Project Intern Xeratic//UU Nomor 6 Tahun 2023.pdf"
23
24 # Format the text from the PDF with line breaks after each point
25 text_with_breaks = format_text_with_line_breaks(pdf_path_with_breaks)
26
27 # Save the formatted text to a new file with line breaks
28 output_with_breaks_path = "C://Users//Ajeng//Documents//Project Intern Xeratic//cipta kerja text"
29
30 with open(output_with_breaks_path, "w") as file:
31     file.write(text_with_breaks)
32
33 output_with_breaks_path
```

4. Lalu block semua script dan run script dengan tekan shift + enter.
5. Hasil file text yang telah diubah akan muncul otomatis di folder yang diinginkan sesuai path.



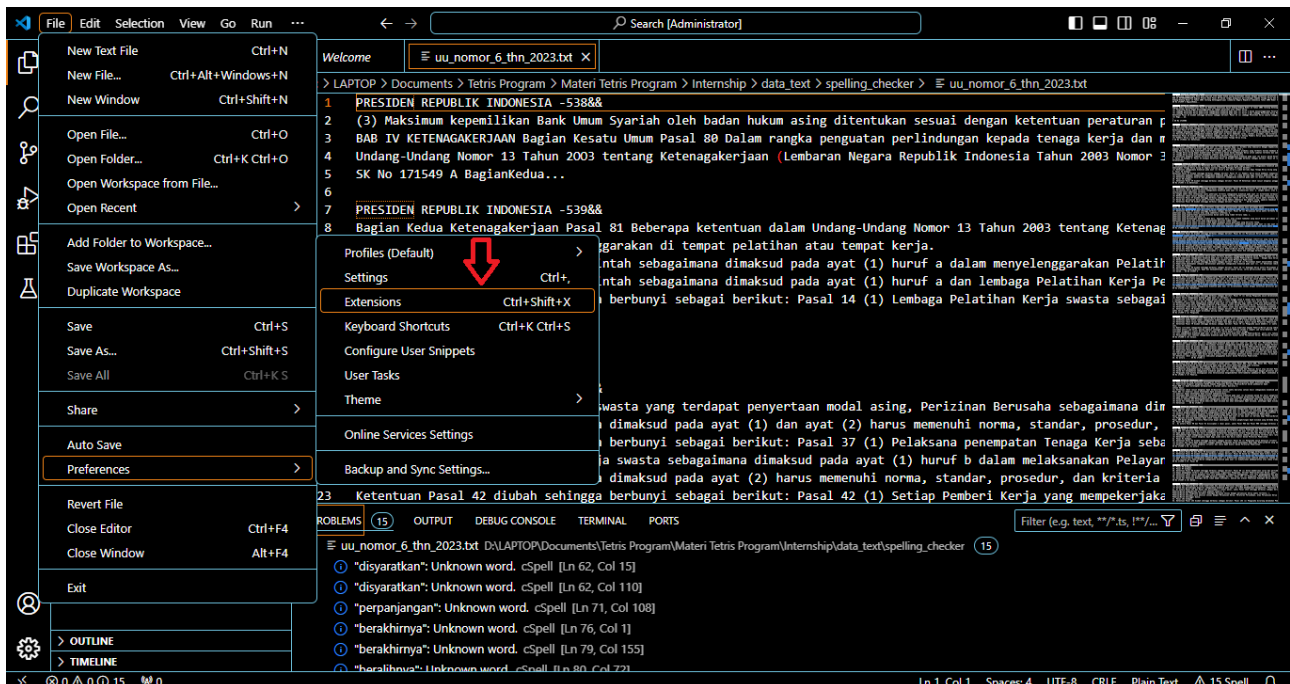


B. Penggunaan Spelling Checker dengan Ekstensi Indonesian – Code Spell Checker

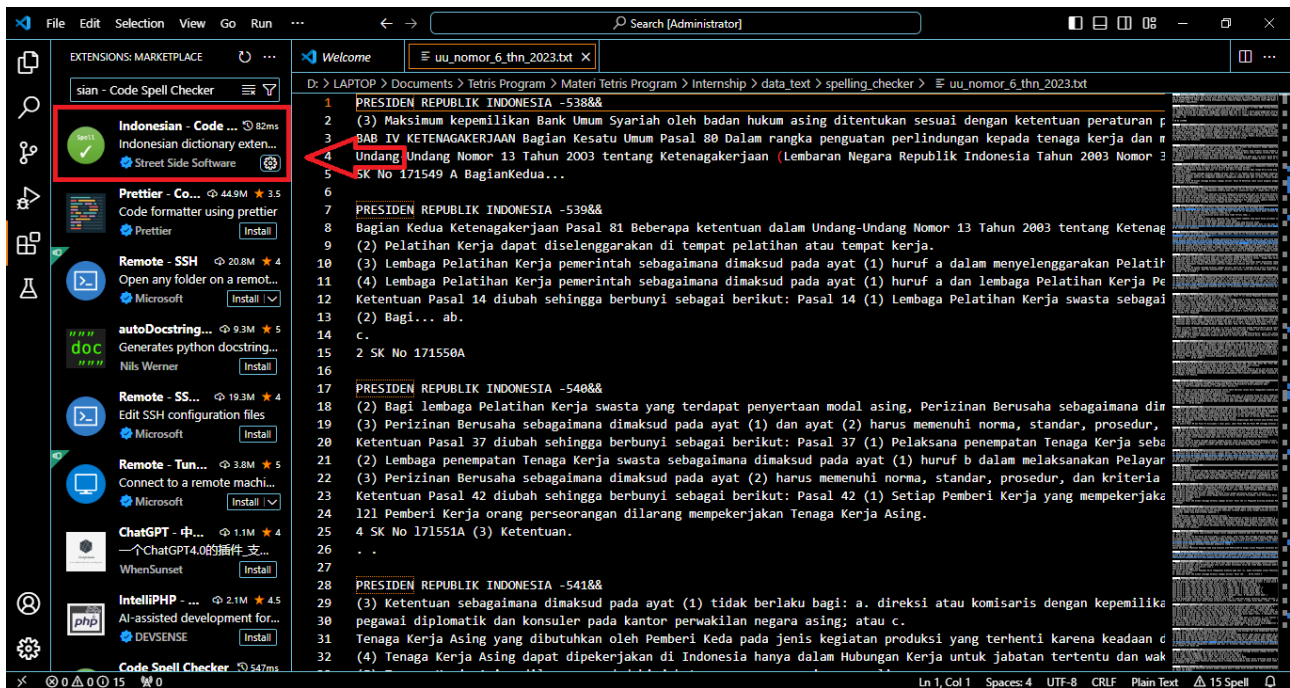
1. Instal Ekstensi Indonesian - Code Spell Checker

a) Buka Marketplace Ekstensi:

- Jalankan Visual Studio Code.



- Buka File, kemudian klik Preferences, pilih Extension.
- Di bagian *Search in Marketplace*, ketik "Indonesian - Code Spell Checker".
- Cari ekstensi buatan "Eko Prasetyo" dan klik Install.



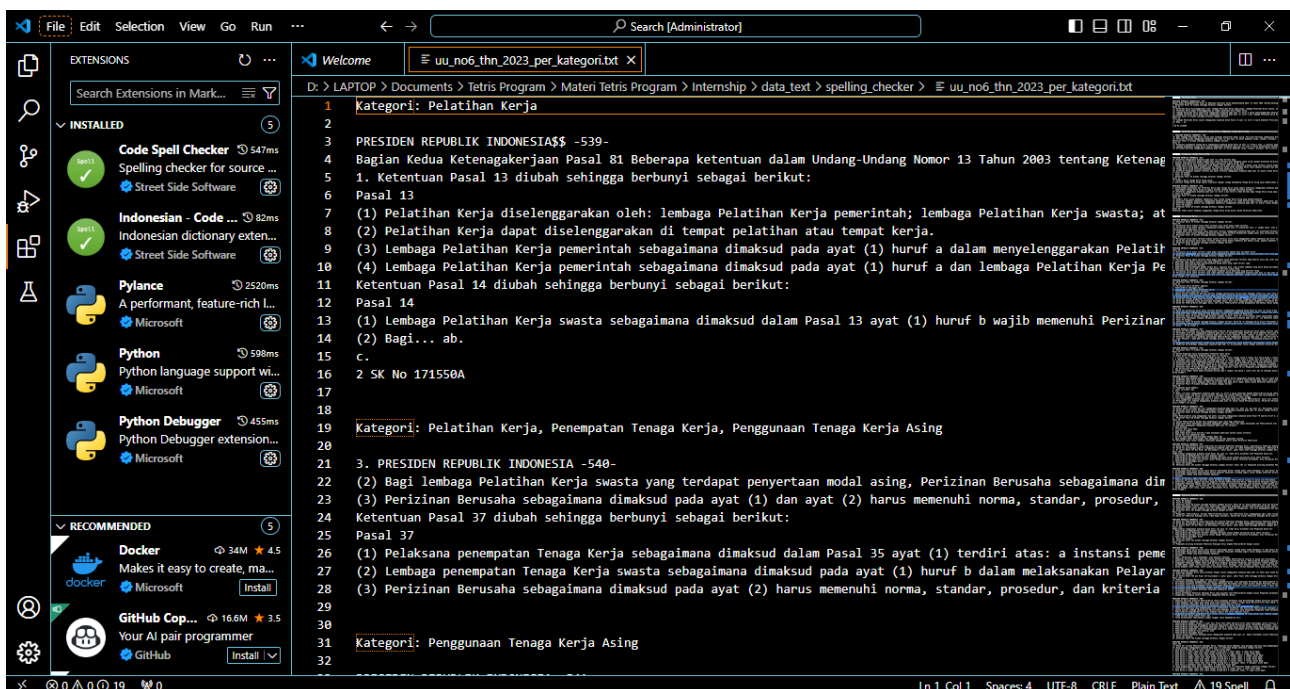
b) Restart Visual Studio Code:

- o Setelah instalasi selesai, tutup dan buka ulang Visual Studio Code.

2. Periksa dan Perbaiki Bahasa Indonesia

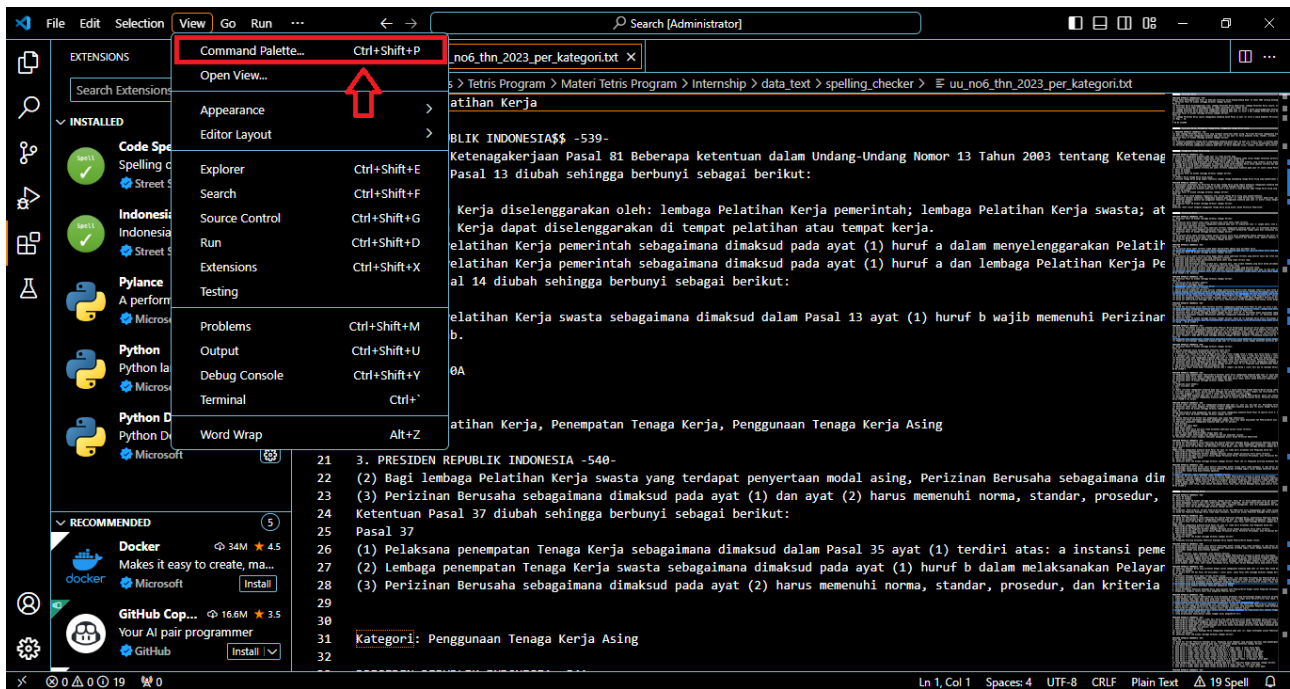
a) Buka File di Visual Studio Code:

- o Arahkan ke direktori yang berisi file teks yang ingin diperiksa.
- o Klik dua kali file tersebut untuk membukanya di Visual Studio Code.

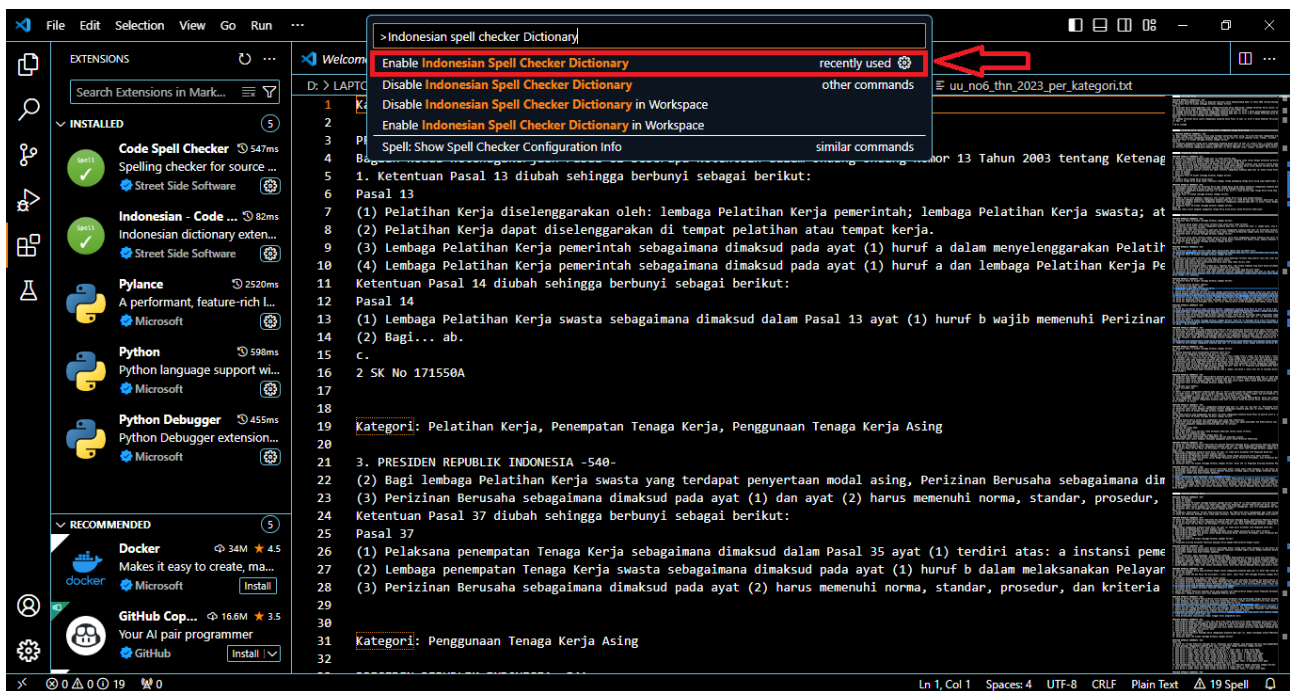


b) Jalankan Indonesian - Code Spell Checker:

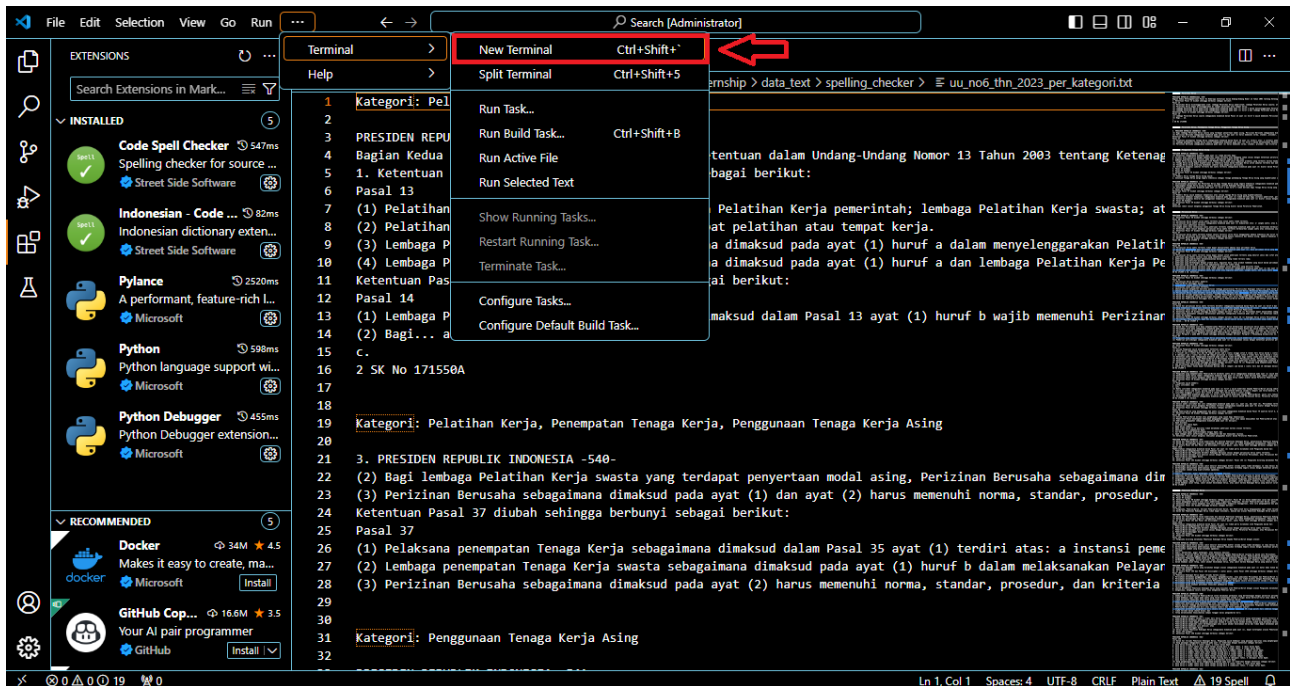
- o Buka Command Palette dengan menekan Ctrl+Shift+P (Windows) atau Cmd+Shift+P (Mac).



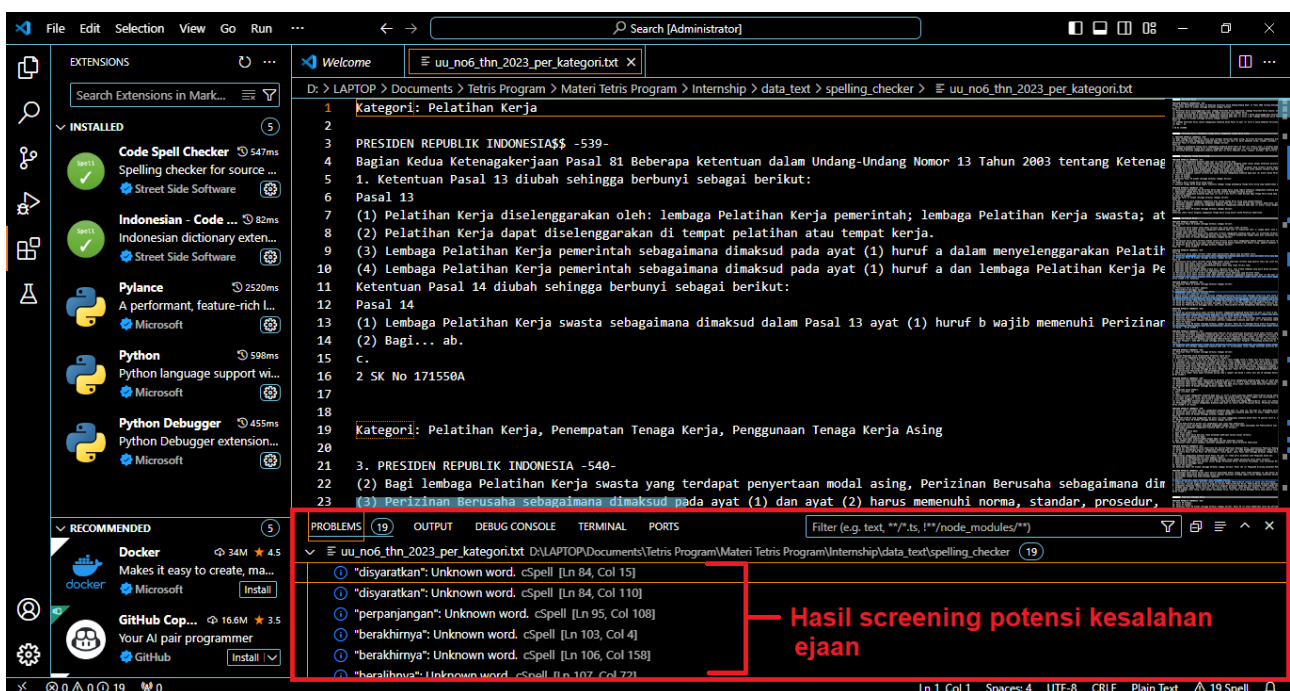
- o Ketik "Enable Indonesian - Code Spell Checker" dan pilih perintah tersebut.



- o Buka New Terminal dengan pilih ... (3 titik di sebelah Run), kemudian pilih Terminal dan klik New Terminal.



- o Ekstensi akan memindai seluruh file dan menyorot potensi kesalahan ejaan dan tata Bahasa.

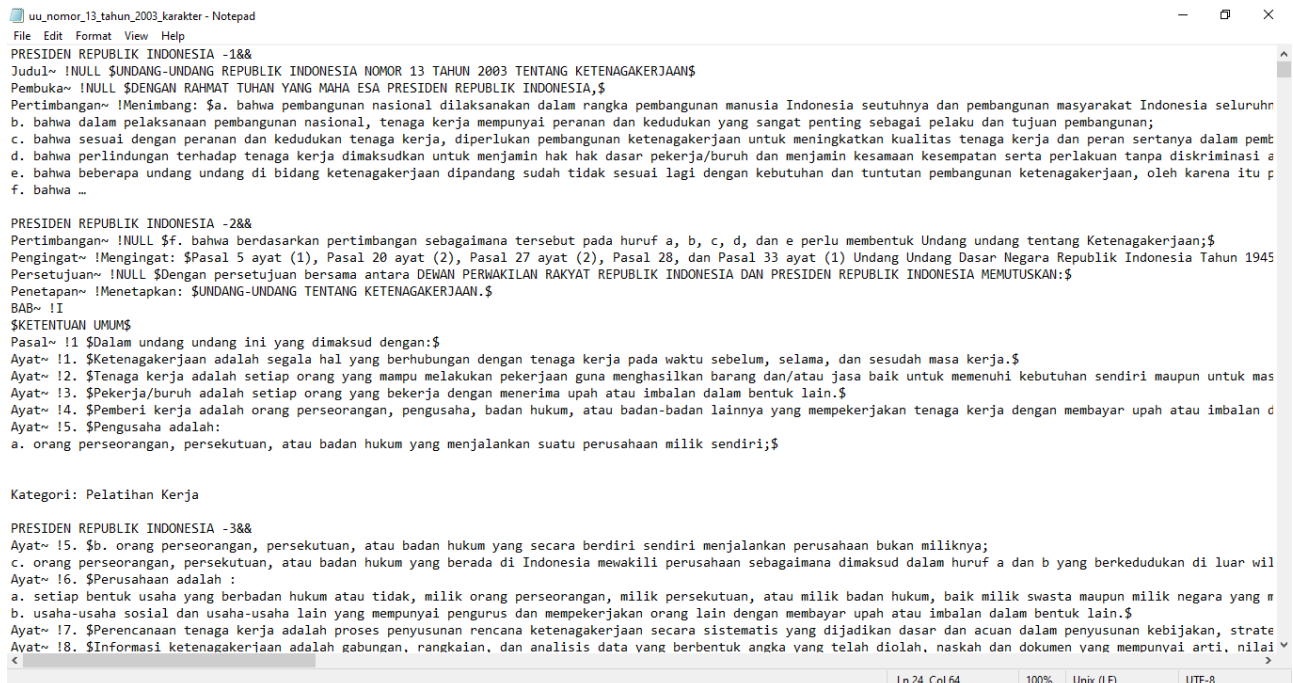


- o Ubah ejaan yang salah menjadi ejaan yang benar secara manual

C. Penggunaan Script untuk Parsing Text Data to Tabular Data

1. Beri karakter pada data teks yang akan diubah ke bentuk tabel dengan ketentuan sebagai berikut:
 - a. Isi kolom Halaman ditandai dengan karakter – (strip) sebelum angka halaman dan karakter && di akhir angka halaman
contoh: -11&& artinya halaman 11
 - b. Isi kolom Context ditandai tanda ~ diakhir kata: Judul, Pembuka, Penetapan, Peningkat, Persetujuan, Pertimbangan, BAB, Bagian, Pasal, Ayat
contoh: Judul~ artinya Judul, Pembuka~ artinya Pembuka
 - c. Isi Kolom Context_number (berupa angka atau huruf) diawali tanda ! sebelum huruf atau angka: huruf BAB, Angka Bagian, Angka Pasal, Angka ayat, dan tambahkan !NULL untuk Judul, Pembuka, Penetapan, Peningkat, Persetujuan, Pertimbangan
Contoh: “Judul~ !NULL”, “Pembuka~ !NULL”, “BAB~ !II”
 - d. Isi Kolom Content ditandai dengan diawali dan diakhiri tanda karakter \$: Isi Judul, Isi Pembuka, Isi Bab, Isi Bagian, Isi Pasal, Isi Ayat, dll.
Contoh: \$UNDANG-UNDANG REPUBLIK INDONESIA NOMOR 13 TAHUN 2003 TENTANG KETENAGAKERJAAN\$

Hasil pemberian karakter secara manual pada data teks akan menjadi seperti berikut:



```
uu_nomor_13_tahun_2003_karakter - Notepad
File Edit Format View Help
PRESIDEN REPUBLIK INDONESIA -1&&
Judul~ !NULL $UNDANG-UNDANG REPUBLIK INDONESIA NOMOR 13 TAHUN 2003 TENTANG KETENAGAKERJAAN$
Pembuka~ !NULL $DENGAN RAHMAT TUHAN YANG MAHA ESA PRESIDEN REPUBLIK INDONESIA,$
Pertimbangan~ !Menimbang: $a. bahwa pembangunan nasional dilaksanakan dalam rangka pembangunan manusia Indonesia seutuhnya dan pembangunan masyarakat Indonesia seluruhn
b. bahwa dalam pelaksanaan pembangunan nasional, tenaga kerja mempunyai peranan dan kedudukan yang sangat penting sebagai pelaku dan tujuan pembangunan;
c. bahwa sesuai dengan peranan dan kedudukan tenaga kerja, diperlukan pembangunan ketenagakerjaan untuk meningkatkan kualitas tenaga kerja dan peran sertanya dalam pemb
d. bahwa perlindungan terhadap tenaga kerja dimaksudkan untuk menjamin hak hak dasar pekerja/buruh dan menjamin kesamaan kesempatan serta perlakuan tanpa diskriminasi a
e. bahwa beberapa undang undang di bidang ketenagakerjaan dipandang sudah tidak sesuai lagi dengan kebutuhan dan tuntutan pembangunan ketenagakerjaan, oleh karena itu p
f. bahwa ...

PRESIDEN REPUBLIK INDONESIA -2&&
Pertimbangan~ !NULL $f. bahwa berdasarkan pertimbangan sebagaimana tersebut pada huruf a, b, c, d, dan e perlu membentuk Undang undang tentang Ketenagakerjaan;$
Peningkat~ !Meningkat: $Pasal 5 ayat (1), Pasal 20 ayat (2), Pasal 27 ayat (2), Pasal 28, dan Pasal 33 ayat (1) Undang Undang Dasar Negara Republik Indonesia Tahun 1945
Persetujuan~ !NULL $Dengan persetujuan bersama antara DEWAN PERAKILAN RAKYAT REPUBLIK INDONESIA DAN PRESIDEN REPUBLIK INDONESIA MEMUTUSKAN:$
Penetapan~ !Menetapkan: $UNDANG-UNDANG TENTANG KETENAGAKERJAAN.$
BAB~ !I
$KETENTUAN UMUM$
Pasal~ !1 $Dalam undang undang ini yang dimaksud dengan:$
Ayat~ !1. $Ketenagakerjaan adalah segala hal yang berhubungan dengan tenaga kerja pada waktu sebelum, selama, dan sesudah masa kerja.$
Ayat~ !2. $Tenaga kerja adalah setiap orang yang mampu melakukan pekerjaan guna menghasilkan barang dan/atau jasa baik untuk memenuhi kebutuhan sendiri maupun untuk mas
Ayat~ !3. $Pekerja/buruh adalah setiap orang yang bekerja dengan menerima upah atau imbalan dalam bentuk lain.$
Ayat~ !4. $Pemberi kerja adalah orang perseorangan, pengusaha, badan hukum, atau badan-badan lainnya yang mempekerjakan tenaga kerja dengan membayar upah atau imbalan d
Ayat~ !5. $Pengusaha adalah:
a. orang perseorangan, persekutuan, atau badan hukum yang menjalankan suatu perusahaan milik sendiri;$
b. orang perseorangan, persekutuan, atau badan hukum yang secara berdiri sendiri menjalankan perusahaan bukan miliknya;
c. orang perseorangan, persekutuan, atau badan hukum yang berada di Indonesia mewakili perusahaan sebagaimana dimaksud dalam huruf a dan b yang berkedudukan di luar wil
Ayat~ !6. $Perusahaan adalah :
a. setiap bentuk usaha yang berbadan hukum atau tidak, milik orang perseorangan, milik persekutuan, atau milik badan hukum, baik milik swasta maupun milik negara yang m
b. usaha-usaha sosial dan usaha-usaha lain yang mempunyai pengurus dan mempekerjakan orang lain dengan membayar upah atau imbalan dalam bentuk lain.$
Ayat~ !7. $Perencanaan tenaga kerja adalah proses penyusunan rencana ketenagakerjaan secara sistematis yang dijadikan dasar dan acuan dalam penyusunan kebijakan, strate
Ayat~ !8. $Informasi ketenagakerjaan adalah gabungan, rangkaian, dan analisis data yang berbentuk angka yang telah diolah, naskah dan dokumen yang mempunyai arti, nilai
<
Ln 24, Col 64 100% Unix (LF) UTF-8
```

2. Buka Script `uu_nomor_13_thn_2003_context.py` untuk file `uu_nomor_13_tahun_2003_karakter.txt` atau `context parsing.py` untuk file `uu_nomor_6_thn_2023_karakter1.txt`. Script ini digunakan untuk mengubah data teks menjadi data tabular dengan berisi kolom halaman, kolom context dan kolom context_number

```
1 import re
2 import pandas as pd
3
4 # Baca isi file teks
5 with open("D:/LAPTOP/Documents/Tetris Program/Materi Tetris Program/Internship/data_text/spelling_checker/uu_nom
6     text = file.read()
7
8 # Proses teks untuk mendapatkan halaman dan konteks
9 matches = re.findall(r"PRESIDEN REPUBLIK INDONESIA -(\d+)&&\n(?:\n(?=PRESIDEN REPUBLIK INDONESIA|\Z)", text, r
10
11 # Susun data menjadi tabel
12 data = []
13 for match in matches:
14     page = match[0]
15     contexts = re.findall(r"(Judul|Pembuka|Pertimbangan|Peningkat|Persetujuan|Penetapan|BAB|Bagian|Paragraf|Pase
16     for context in contexts:
17         data.append((page, context[0], context[1]))
18     #else:
19         #data.append((page, "(NULL)", "(NULL)"))
20
21 # Buat dataframe menggunakan pandas
22 df = pd.DataFrame(data, columns=["halaman", "context", "context_number"])
23
24 # Hapus karakter && dan ~ dari kolom halaman dan context
25 df['halaman'] = df['halaman'].str.replace("&", "")
26 df['context'] = df['context'].str.replace("~", "")
27 df['context_number'] = df['context_number'].str.replace("!", "")
28
29 # Tampilkan dataframe
30 print(df)
31
32 # Simpan dataframe ke file Excel
```

3. Ubah path sesuai folder dimana file text berkarakter yang akan diubah ke data tabel berada pada script (line 5):
- with open("D:/LAPTOP/Documents/Tetris Program/Materi Tetris Program/Internship/data_text/spelling_checker/ uu_nomor_13_tahun_2003_karakter.txt","r") as file:

```
1 import re
2 import pandas as pd
3
4 # Baca isi file teks
5 with open("D:/LAPTOP/Documents/Tetris Program/Materi Tetris Program/Internship/data_text/spelling_checker/uu_nom
6     text = file.read()
7
8 # Proses teks untuk mendapatkan halaman dan konteks
9 matches = re.findall(r"PRESIDEN REPUBLIK INDONESIA -(\d+)&&\n(?:\n(?=PRESIDEN REPUBLIK INDONESIA|\Z)", text, r
10
11 # Susun data menjadi tabel
12 data = []
13 for match in matches:
14     page = match[0]
15     contexts = re.findall(r"(Judul|Pembuka|Pertimbangan|Peningkat|Persetujuan|Penetapan|BAB|Bagian|Paragraf|Pase
16     for context in contexts:
17         data.append((page, context[0], context[1]))
18     #else:
19         #data.append((page, "(NULL)", "(NULL)"))
20
21 # Buat dataframe menggunakan pandas
22 df = pd.DataFrame(data, columns=["halaman", "context", "context_number"])
23
24 # Hapus karakter && dan ~ dari kolom halaman dan context
25 df['halaman'] = df['halaman'].str.replace("&", "")
26 df['context'] = df['context'].str.replace("~", "")
27 df['context_number'] = df['context_number'].str.replace("!", "")
28
29 # Tampilkan dataframe
30 print(df)
31
32 # Simpan dataframe ke file Excel
```

4. Ubah path sesuai folder dimana data tabel yang telah diubah akan ditempatkan pada script (line 33):
- df.to_excel("D:/LAPTOP/Documents/Tetris Program/Materi Tetris Program/Internship/data_table/tabular/uu_nomor_13_tahun_2003_context.xlsx", index=False)

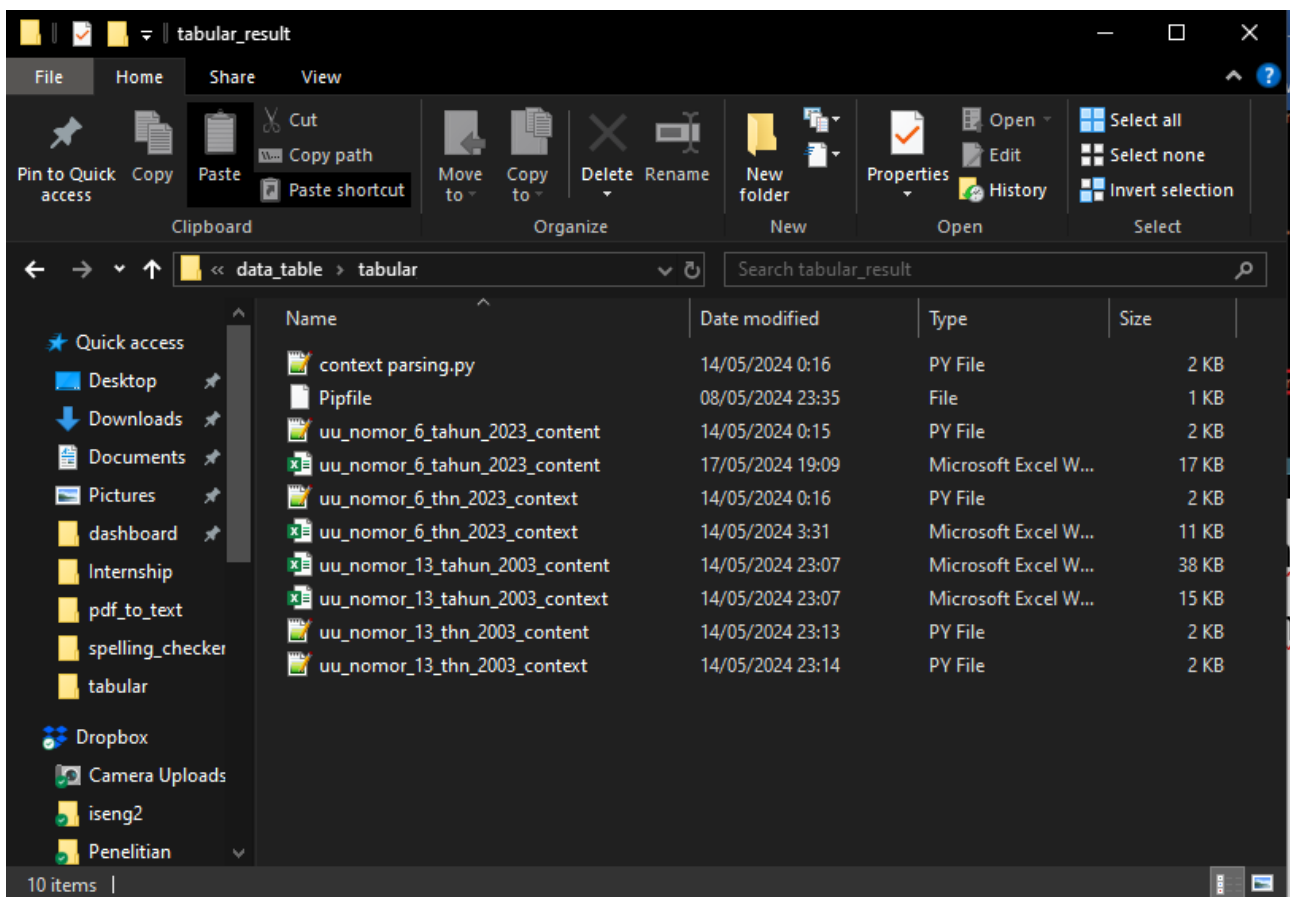

```

5 with open("D:/LAPTOP/Documents/Tetris Program/Materi Tetris Program/Internship/data_table/tabular/uu_nomor_13_thn_2003_context.py") as f:
6     text = f.read()
7
8 # Proses teks untuk mendapatkan halaman dan konteks
9 matches = re.findall(r"PRESIDEN REPUBLIK INDONESIA -(\d+)&\n(?:\n(?=PRESIDEN REPUBLIK INDONESIA)|\Z)", text, re.M)
10
11 # Susun data menjadi tabel
12 data = []
13 for match in matches:
14     page = match[0]
15     contexts = re.findall(r"(Judul|Pembuka|Pertimbangan|Peningkat|Persetujuan|Penetapan|BAB|Bagian|Paragraf|Pasa",
16     for context in contexts:
17         data.append((page, context[0], context[1]))
18     #else:
19         #data.append((page, "(NULL)", "(NULL)"))
20
21 # Buat dataframe menggunakan pandas
22 df = pd.DataFrame(data, columns=["halaman", "context", "context_number"])
23
24 # Hapus karakter & dan ~ dari kolom halaman dan context
25 df['halaman'] = df['halaman'].str.replace("&", "")
26 df['context'] = df['context'].str.replace("~", "")
27 df['context_number'] = df['context_number'].str.replace("!", "")
28
29 # Tampilkan dataframe
30 print(df)
31
32 # Simpan dataframe ke file Excel
33 df.to_excel("D:/LAPTOP/Documents/Tetris Program/Materi Tetris Program/Internship/data table/tabular/uu nomor 13")
34

```

path untuk data tabel

5. Lalu block semua script dan run script dengan tekan shift + enter.
6. Hasil data tabel yang telah diubah akan muncul otomatis di folder yang diinginkan sesuai path.



7. Cara yang sama juga pada script file *content parsing.py*. File script ini dapat digunakan pada kedua data text, yaitu file *uu_nomor_13_tahun_2003_karakter.txt* dan file *uu_nomor_6_thn_2023_karakter1.txt*. Script ini digunakan untuk mengubah data teks menjadi data tabular dengan berisi kolom halaman dan kolom content.

Hasil data tabel akan seperti berikut ini:

Data tabel *uu_nomor_13_tahun_2003_context.xlsx* dan *uu_nomor_13_tahun_2003_content.xlsx*

halaman	context	text_number
1	Judul	NULL
1	Pembuka	NULL
1	Pertimbar Menimbang	
2	Pertimbar	NULL
2	Pengingat Mengingat	
2	Persetujuan	NULL
2	Penetapan Menetapkan	
2	BAB	1
2	Pasal	1
2	Ayat	1
2	Ayat	2
2	Ayat	3
2	Ayat	4
2	Ayat	5
2	Ayat	6
2	Ayat	7
2	Ayat	8
2	Ayat	9
2	Ayat	10
2	Ayat	11
2	Ayat	12

halaman	content
1	UNDANG-UNDANG REPUBLIK INDONESIA NOMOR 13 TAHUN 2003 TENTANG KETENAGAKERJAAN
1	DENGAN RAHMAT TUHAN YANG MAHA ESA PRESIDEN REPUBLIK INDONESIA,
1	a. bahwa pembangunan nasional dilaksanakan dalam rangka pembangunan manusia Indonesia
2	f. bahwa berdasarkan pertimbangan sebagaimana tersebut pada huruf a, b, c, d, dan e perlu
2	Pasal 5 ayat (1), Pasal 20 ayat (2), Pasal 27 ayat (2), Pasal 28, dan Pasal 33 ayat (1) Undang Und
2	Dengan persetujuan bersama antara DEWAN PERWAKILAN RAKYAT REPUBLIK INDONESIA DA
2	UNDANG-UNDANG TENTANG KETENAGAKERJAAN.
2	KETENTUAN UMUM
2	Dalam undang undang ini yang dimaksud dengan:
2	Ketenagakerjaan adalah segala hal yang berhubungan dengan tenaga kerja pada waktu sebel
2	Tenaga kerja adalah setiap orang yang mampu melakukan pekerjaan guna menghasilkan bar
2	Pekerja/buruh adalah setiap orang yang bekerja dengan menerima upah atau imbalan dalam
2	Pemberi kerja adalah orang perseorangan, pengusaha, badan hukum, atau badan-badan lain
2	Pengusaha adalah:a. orang perseorangan, persekutuan, atau badan hukum yang menjalankan
2	b. orang perseorangan, persekutuan, atau badan hukum yang secara berdiri sendiri menjalar
2	Perusahaan adalah :a. setiap bentuk usaha yang berbadan hukum atau tidak, milik orang pers
2	Perencanaan tenaga kerja adalah proses penyusunan rencana ketenagakerjaan secara sistem
2	Informasi ketenagakerjaan adalah gabungan, rangkaian, dan analisis data yang berbentuk an
2	Pelatihan kerja adalah keseluruhan kegiatan untuk memberi, memperoleh, meningkatkan, s
2	Kompetensi kerja adalah kemampuan kerja setiap individu yang mencakup aspek pengetahu
2	Pemagangan adalah bagian dari sistem pelatihan kerja yang diselenggarakan secara terpadu
2	Delapanan penempatan tenaga kerja adalah kegiatan untuk memmertemukan tenaga kerja di

Data tabel *uu_nomor_6_tahun_2023_context.xlsx* dan *uu_nomor_6_tahun_2023_content.xlsx*

halaman	context	text_number
538	BAB	IV
538	Bagian	Kesatu
538	Pasal	80
539	Bagian	Kedua
539	Pasal	81
539	Ayat	1
539	Ayat	2
540	Ayat	2
540	Ayat	3
540	Ayat	4
541	Ayat	4
541	Ayat	5
541	Ayat	6
541	Ayat	7
541	Ayat	7
542	Ayat	7
542	Ayat	8
542	Ayat	9
542	Ayat	10
542	Ayat	11
543	Ayat	12
543	Ayat	13
543	Ayat	14

halaman	content
538	Dalam rangka penguatan perlindungan kepada tenaga kerja dan meningkatkan peran dan ke
539	Beberapa ketentuan dalam Undang-Undang Nomor 13 Tahun 2003 tentang Ketenagakerjaan
539	Ketentuan Pasal 13 diubah sehingga berbunyi sebagai berikut: Pasal 13 (1) Pelatihan Kerja di
539	Ketentuan Pasal 14 diubah sehingga berbunyi sebagai berikut: Pasal 14 (1) Lembaga Pelatihar
540	(2) Bagi lembaga Pelatihan Kerja swasta yang terdapat penyertaan modal asing, Perizinan Be
540	Ketentuan Pasal 37 diubah sehingga berbunyi sebagai berikut: Pasal 37 (1) Pelaksana penem
540	Ketentuan Pasal 42 diubah sehingga berbunyi sebagai berikut: Pasal 42 (1) Setiap Pemberi Ke
541	(3) Ketentuan sebagaimana dimaksud pada ayat (1) tidak berlaku bagi: a. direksi atau komis
541	Pasal 43 dihapus.
541	Pasal 44 dihapus.
541	Ketentuan Pasal 45 diubah sehingga berbunyi sebagai berikut: Pasal 45 (1) Pemberi Kerja Ter
542	(1) b. melaksanakan pendidikan dan Pelatihan Kerja bagi Tenaga Kerja warga negara Indones
542	Pasal 46 dihapus.
542	Ketentuan Pasal 47 diubah sehingga berbunyi sebagai berikut: Pasal 47 (1) Pemberi Kerja wa
542	Pasal 48 dihapus.
542	Ketentuan Pasal 49 diubah sehingga berbunyi sebagai berikut: Pasal 49 Ketentuan lebih lanju
543	Ketentuan Pasal 56 diubah sehingga berbunyi sebagai berikut: Pasal 56 (1) Perjanjian Kerja d
543	Ketentuan Pasal 57 diubah sehingga berbunyi sebagai berikut: Pasal 57 (1) Perjanjian kerja w
543	Ketentuan Pasal 58 diubah sehingga berbunyi sebagai berikut: Pasal 58
544	Pasal 58 (1) Perjanjian kerja waktu tertentu tidak dapat mensyaratkan adanya masa percobaan
544	Ketentuan Pasal 59 diubah sehingga berbunyi sebagai berikut: Pasal 59 (1) Perjanjian kerja w
545	Ketentuan Pasal 61 diubah sehingga berbunyi sebagai berikut: Pasal 61 (1) Perjanjian Kerja h

D. Menggabung 2 Data Tabel dengan Manipulation Script

1. Ubah path sesuai folder dimana 2 data tabel *uu_nomor_13_tahun_2003_context.xlsx* dan *uu_nomor_13_tahun_2003_content.xlsx* yang akan digabungkan berada pada script *manipulation script uu_13.ipynb*:

File Edit Selection View Go Run ... Search [Administrator]

manipulation script uu_13.ipynb manipulation script uu_06.ipynb

Python 3.12.1

Data Manipulation Script

```
import pandas as pd
import numpy as np

content= pd.read_excel("C://Users//Ajeng//Documents//Project Intern Xeratic//final project repo//uu_nomor_13_tahun_2003_cont
context = pd.read_excel("C://Users//Ajeng//Documents//Project Intern Xeratic//final project repo//uu_nomor_13_tahun_2003_cont

data = pd.concat([content,context], axis=1)
display(data)
```

path untuk file content

path untuk file context

halaman1	content	halaman	context	context_number	
0	1	UNDANG-UNDANG REPUBLIK INDONESIA NOMOR 13 TAHUN...	1	Judul	NaN
1	1	DENGAN RAHMAT TUHAN YANG MAHA ESA PRESIDEN REP...	1	Pembuka	NaN

PROBLEMS (22) OUTPUT DEBUG CONSOLE TERMINAL PORTS JUPYTER

numpy 1.26.4
openpyxl 3.1.2
packaging 23.2
pandas 2.2.0
pandocfilters 1.5.1
parso 0.8.3
pickleshare 0.7.5
pillow 10.2.0

Python
Python
Python
Python
Python
powershell

Ln 2, Col 1 Spaces: 4 Cell 2 of 14 4 Spell

2. klik run untuk menampilkan hasilnya

File Edit Selection View Go Run ... Search [Administrator]

manipulation script uu_13.ipynb manipulation script uu_06.ipynb

Python 3.12.1

halaman1	content	halaman	context	context_number	
0	1	UNDANG-UNDANG REPUBLIK INDONESIA NOMOR 13 TAHUN...	1	Judul	NaN
1	1	DENGAN RAHMAT TUHAN YANG MAHA ESA PRESIDEN REP...	1	Pembuka	NaN
2	1	a. bahwa pembangunan nasional dilaksanakan dal...	1	Pertimbangan	Menimbang
3	2	f. bahwa berdasarkan pertimbangan sebagaimana ...	2	Pertimbangan	NaN
4	2	Pasal 5 ayat (1), Pasal 20 ayat (2), Pasal 27 ...	2	Pengingat	Mengingat
...
688	79	Undang-undang Nomor 14 Tahun 1969 tentang Kete...	79	Ayat	12
689	79	Undang-undang Nomor 25 Tahun 1997 tentang Kete...	79	Ayat	13
690	79	Undang-undang Nomor 11 Tahun 1998 tentang Peru...	79	Ayat	14
691	79	Undang-undang Nomor 28 Tahun 2000 tentang Pene...	79	Ayat	15
692	79	Undang-undang ini mulai berlaku pada tanggal d...	79	Pasal	193

693 rows x 5 columns

hasil running

PROBLEMS (22) OUTPUT DEBUG CONSOLE TERMINAL PORTS JUPYTER

numpy 1.26.4
openpyxl 3.1.2
packaging 23.2
pandas 2.2.0
pandocfilters 1.5.1
parso 0.8.3
pickleshare 0.7.5
pillow 10.2.0

Python
Python
Python
Python
Python
powershell

Ln 2, Col 1 Spaces: 4 Cell 2 of 14 4 Spell

3. klik run pada script selanjutnya dan menampilkan hasilnya:

The screenshot shows a Jupyter Notebook interface with the following components:

- EXPLORER:** Shows the file structure with 'manipulation script uu_13.ipynb' selected.
- Code Cell:** Contains the following Python code:

```
data = data.drop(data.columns[0], axis=1)
data.head()
```
- Output:** Displays the first 5 rows of the data table:

	content	halaman	context	context_number
0	UNDANG-UNDANG REPUBLIK INDONESIA NOMOR 13 TAHUN...	1	Judul	NaN
1	DENGAN RAHMAT TUHAN YANG MAHA ESA PRESIDEN REP...	1	Pembuka	NaN
2	a. bahwa pembangunan nasional dilaksanakan dal...	1	Pertimbangan	Menimbang
3	f. bahwa berdasarkan pertimbangan sebagaimana ...	2	Pertimbangan	NaN
4	Pasal 5 ayat (1), Pasal 20 ayat (2), Pasal 27 ...	2	Pengingat	Mengingat
- TERMINAL:** Shows the output of the script, including the number of rows and columns (693 rows x 5 columns).

4. Ubah path sesuai folder dimana hasil penggabungan 2 data tabel

uu_nomor_13_tahun_2003_context.xlsx dan *uu_nomor_13_tahun_2003_content.xlsx* diletakkan.

The screenshot shows a Jupyter Notebook interface with the following components:

- EXPLORER:** Shows the file structure with 'manipulation script uu_13.ipynb' selected.
- Code Cell:** Contains the following Python code:

```
#data.to_excel('C://Users/Ajeng/Documents/Project Intern Xeratic/final project repo//merge_uu_13.xlsx')

generate pasal

# Identify the rows containing 'pasal' in the 'context' column
pasal_indices = data[data['context'].str.contains('pasal', case=False, na=False)].index

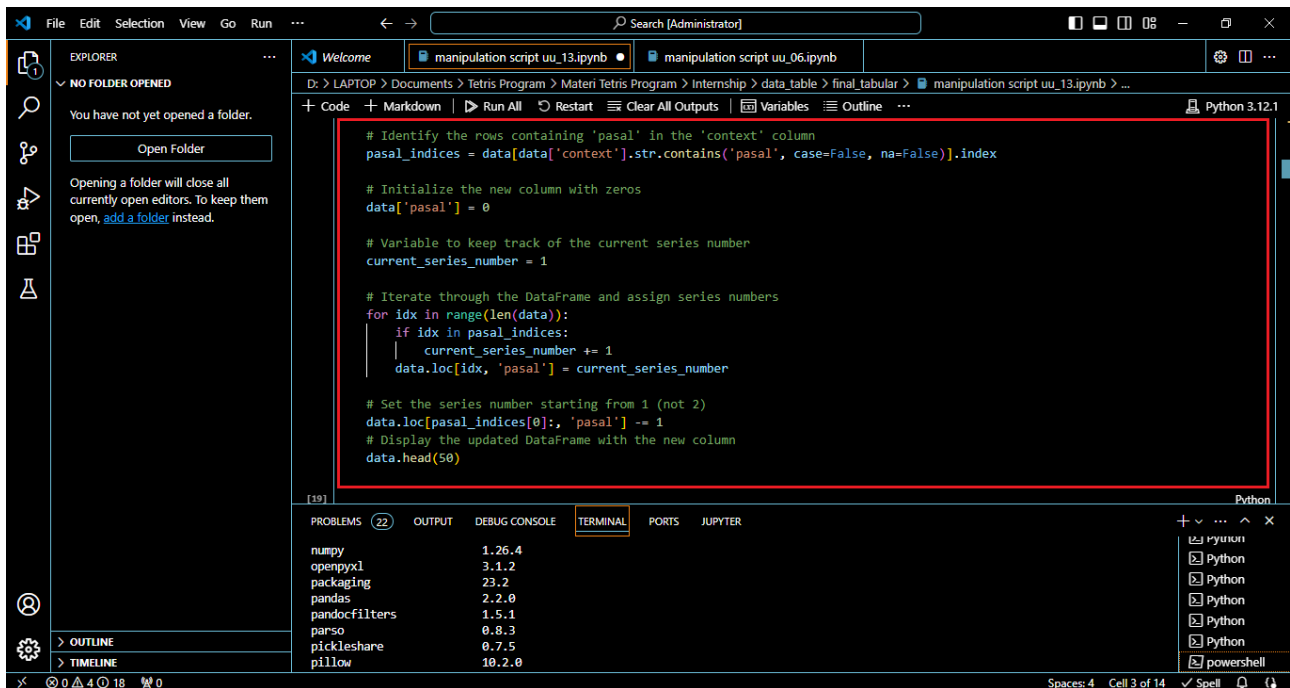
# Initialize the new column with zeros
data['pasal'] = 0

# Variable to keep track of the current series number
current_series_number = 1

# Iterate through the DataFrame and assign series numbers
for idx in range(len(data)):
    if idx in pasal_indices:
        current_series_number += 1
```
- Output:** Displays the first 5 rows of the data table, identical to the one in the previous screenshot.
- TERMINAL:** Shows the output of the script, including the number of rows and columns (693 rows x 5 columns).

A red arrow points to the file path in the code cell, labeled "path hasil penggabungan tabel".

5. klik run pada script selanjutnya untuk menampilkan hasil generate pasal



The screenshot shows a Jupyter Notebook interface with a Python script in the main cell. The script identifies rows containing 'pasal' in the 'context' column, initializes a new column 'pasal' with zeros, and iterates through the DataFrame to assign series numbers. The script is as follows:

```
# Identify the rows containing 'pasal' in the 'context' column
pasal_indices = data[data['context'].str.contains('pasal', case=False, na=False)].index

# Initialize the new column with zeros
data['pasal'] = 0

# Variable to keep track of the current series number
current_series_number = 1


# Iterate through the DataFrame and assign series numbers
for idx in range(len(data)):
    if idx in pasal_indices:
        current_series_number += 1
        data.loc[idx, 'pasal'] = current_series_number

# Set the series number starting from 1 (not 2)
data.loc[pasal_indices[0], 'pasal'] -= 1
# Display the updated DataFrame with the new column
data.head(50)
```

The bottom of the interface shows the 'TERMINAL' tab with a list of installed packages and their versions:

Package	Version
numpy	1.26.4
openpyxl	3.1.2
packaging	23.2
pandas	2.2.0
pandocfilters	1.5.1
parso	0.8.3
pickleshare	0.7.5
pillow	10.2.0

6. hasil running akan terlihat seperti berikut ini:



The screenshot shows the Jupyter Notebook interface with the output of the script displayed in the 'TERMINAL' tab. The output is a table with 5 columns: 'content', 'halaman', 'context', 'context_number', and 'pasal'. The table contains 13 rows of data. A red box highlights the table, and the text 'Hasil generate pasal' is written in red to the right of the table.

	content	halaman	context	context_number	pasal
0	UNDANG-UNDANG REPUBLIK INDONESIA NOMOR 13 TAHUN...	1	Judul	NaN	1
1	DENGAN RAHMAT TUHAN YANG MAHA ESA PRESIDEN REP...	1	Pembuka	NaN	1
2	a. bahwa pembangunan nasional dilaksanakan dal...	1	Pertimbangan	Menimbang	1
3	f. bahwa berdasarkan pertimbangan sebagaimana ...	2	Pertimbangan	NaN	1
4	Pasal 5 ayat (1), Pasal 20 ayat (2), Pasal 27 ...	2	Pengingat	Mengingat	1
5	Dengan persetujuan bersama antara DEWAN PERWAK...	2	Persetujuan	NaN	1
6	UNDANG-UNDANG TENTANG KETENAGAKERJAAN.	2	Penetapan	Menetapkan	1
7	KETENTUAN UMUM	2	BAB	I	1
8	Dalam undang undang ini yang dimaksud dengan:	2	Pasal	1	1
9	Ketenagakerjaan adalah segala hal yang berhubu...	2	Ayat	1	1
10	Tenaga kerja adalah setiap orang yang mampu me...	2	Ayat	2	1
11	Pekerja/buruh adalah setiap orang yang bekerja...	2	Ayat	3	1
12	Pemberi kerja adalah orang perseorangan, pengu...	2	Ayat	4	1
13	Pengusaha adalah/na. orang perseorangan, pers...	2	Ayat	5	1

7. semua script di running tahap per tahap hingga selesai

8. untuk script yang terakhir, ubah path sesuai folder dimana hasil akhir penggabungan 2 data tabel yang telah diubah akan ditempatkan pada script yang terakhir.

The screenshot shows a Jupyter Notebook interface with a dark theme. The Explorer panel on the left shows the file structure. The main area displays a Python script with a red arrow pointing to the file path in the code cell. The output panel at the bottom shows the results of the script execution.

Code Cell:

```
data.to_excel('C://Users//Ajeng//Documents//Project Intern Xeratic//final project repo//final_uu_13_03.xlsx')
```

Output:

	numpy	openpyxl	packaging	pandas	pandocfilters	parso	pickleshare	pillow
1.26.4	3.1.2	23.2	2.2.0	1.5.1	0.8.3	0.7.5	10.2.0	

9. hasil akhir penggabungan 2 tabel akan menjadi seperti berikut ini:

The screenshot shows an Excel spreadsheet with the following data:

uu_id	bab	konten_ba	bagian	ama_ba	pasal	context	text_num	content	halaman
0	uu_13_2003					Judul		UNDANG-	1
1	uu_13_2003					Pembuka		DENGAN f	1
2	uu_13_2003					Pertimbar Menimbaia. bahwa j			1
3	uu_13_2003					Pertimbangan f. bahwa t			2
4	uu_13_2003					Pengingat Menginga Pasal 5 ayi			2
5	uu_13_2003					Persetujuan Dengan pi			2
6	uu_13_2003					Penetapan Menetapk UNDANG-			2
7	uu_13_2003	KETENTUAN UMUM				BAB I KETENTUA			2
8	uu_13_2003	KETENTUAN UMUM				1 Pasal 1 Dalam uni			2
9	uu_13_2003	KETENTUAN UMUM				1 Ayat 1 Ketenagal			2
10	uu_13_2003	KETENTUAN UMUM				1 Ayat 2 Tenaga ke			2
11	uu_13_2003	KETENTUAN UMUM				1 Ayat 3 Pekerja/b			2
12	uu_13_2003	KETENTUAN UMUM				1 Ayat 4 Pemberi k			2
13	uu_13_2003	KETENTUAN UMUM				1 Ayat 5 Pengusah			2
14	uu_13_2003	KETENTUAN UMUM				1 Ayat 5 b. orang p			3
15	uu_13_2003	KETENTUAN UMUM				1 Ayat 6 Perusaha			3
16	uu_13_2003	KETENTUAN UMUM				1 Ayat 7 Perencana			3
17	uu_13_2003	KETENTUAN UMUM				1 Ayat 8 Informasi			3
18	uu_13_2003	KETENTUAN UMUM				1 Ayat 9 Pelatihan			3
19	uu_13_2003	KETENTUAN UMUM				1 Ayat 10 Kompet			3
20	uu_13_2003	KETENTUAN UMUM				1 Ayat 11 Pemagan			3
21	uu_13_2003	KETENTUAN UMUM				1 Ayat 12 Pelayanar			4
22	uu_13_2003	KETENTUAN UMUM				1 Ayat 13 Tenaga ke			4
23	uu_13_2003	KETENTUAN UMUM				1 Ayat 14 Perjanjian			4
24	uu_13_2003	KETENTUAN UMUM				1 Ayat 15 Hubungan			4