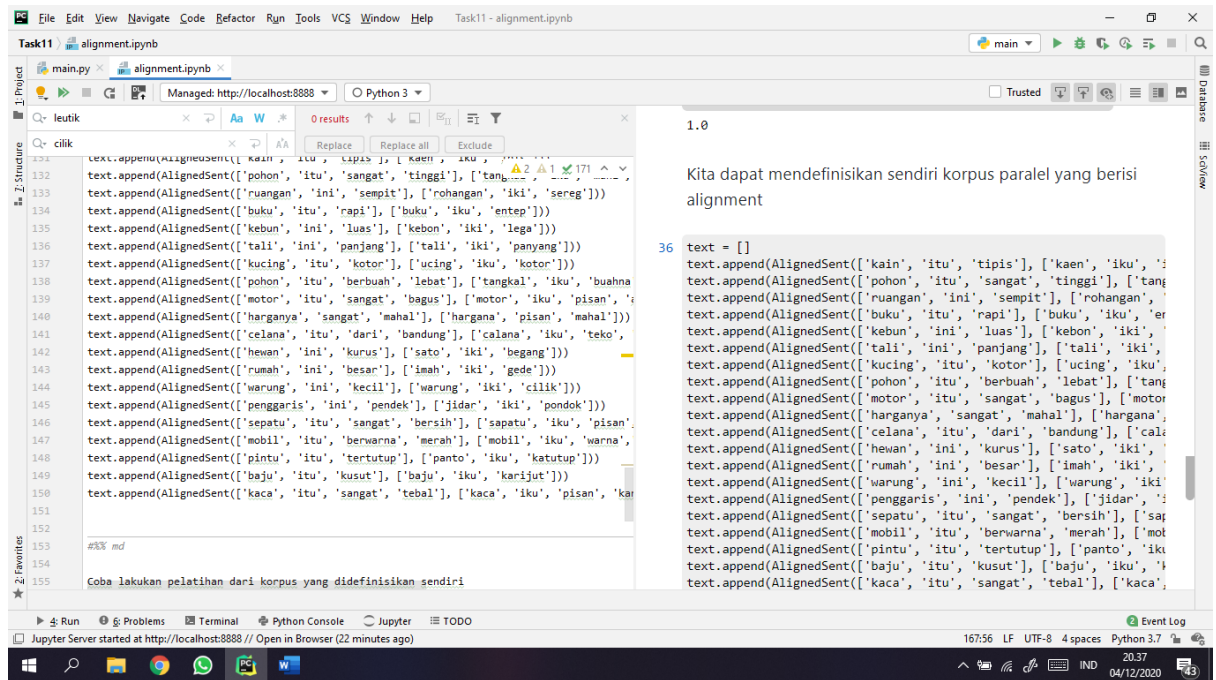


Muhammad Faisal Amir  
1301198497  
IF-41-GAB03

1. Buatlah program untuk pelatihan alignment dari korpus parallel, menggunakan NLTK, dengan algoritma IBMModel1.



```
File Edit View Navigate Code Refactor Run Tools VCS Window Help Task11 - alignment.ipynb
Task11 alignment.ipynb
main
0 results
leutik
cliik
text.append(AlignedSent(['kain', 'itu', 'sangat', 'tinggi'], ['tangan', 'itu', 'sangat', 'tinggi']))
text.append(AlignedSent(['ruangan', 'ini', 'sempit'], ['ruangan', 'iki', 'sempit']))
text.append(AlignedSent(['buku', 'itu', 'rapi'], ['buku', 'iku', 'entep']))
text.append(AlignedSent(['kebun', 'ini', 'luas'], ['kebun', 'iki', 'lega']))
text.append(AlignedSent(['tali', 'ini', 'panjang'], ['tali', 'iki', 'panjang']))
text.append(AlignedSent(['kucing', 'itu', 'kotor'], ['kucing', 'iku', 'kotor']))
text.append(AlignedSent(['pohon', 'itu', 'berbuah', 'lebat'], ['pohon', 'iku', 'berbuah', 'lebat']))
text.append(AlignedSent(['motor', 'itu', 'sangat', 'bagus'], ['motor', 'iku', 'sangat', 'bagus']))
text.append(AlignedSent(['harganya', 'sangat', 'mahal'], ['harganya', 'pisan', 'mahal']))
text.append(AlignedSent(['celana', 'itu', 'dari', 'bandung'], ['celana', 'iku', 'dari', 'teko']))
text.append(AlignedSent(['hewan', 'ini', 'kurus'], ['sato', 'iki', 'kurus', 'begang']))
text.append(AlignedSent(['rumah', 'ini', 'besar'], ['rumah', 'iki', 'gede']))
text.append(AlignedSent(['warung', 'ini', 'kecil'], ['warung', 'iki', 'cilik']))
text.append(AlignedSent(['penggaris', 'ini', 'pendek'], ['penggaris', 'iku', 'pendek']))
text.append(AlignedSent(['sepatu', 'itu', 'sangat', 'bersih'], ['sepatu', 'iku', 'sangat', 'bersih']))
text.append(AlignedSent(['mobil', 'itu', 'berwarna', 'merah'], ['mobil', 'iku', 'berwarna', 'merah']))
text.append(AlignedSent(['pintu', 'itu', 'tertutup'], ['pintu', 'iku', 'tertutup']))
text.append(AlignedSent(['baju', 'itu', 'kusut'], ['baju', 'iku', 'karijut']))
text.append(AlignedSent(['kaca', 'itu', 'sangat', 'tebal'], ['kaca', 'iku', 'pisan', 'kaca']))

# %% md
Coba lakukan pelatihan dari korpus yang didefinisikan sendiri

1.0
Kita dapat mendefinisikan sendiri korpus parallel yang berisi alignment
36 text = []
text.append(AlignedSent(['kain', 'itu', 'tipis'], ['kaen', 'iku', 'tipis']))
text.append(AlignedSent(['pohon', 'itu', 'sangat', 'tinggi'], ['tang', 'iku', 'sangat', 'tinggi']))
text.append(AlignedSent(['ruangan', 'ini', 'sempit'], ['ruangan', 'er', 'sangat', 'sempit']))
text.append(AlignedSent(['buku', 'itu', 'rapi'], ['buku', 'iku', 'er', 'sangat', 'rapi']))
text.append(AlignedSent(['kebun', 'ini', 'luas'], ['kebun', 'iki', 'er', 'sangat', 'luas']))
text.append(AlignedSent(['tali', 'ini', 'panjang'], ['tali', 'iki', 'er', 'sangat', 'panjang']))
text.append(AlignedSent(['kucing', 'itu', 'kotor'], ['kucing', 'iku', 'er', 'sangat', 'kotor']))
text.append(AlignedSent(['pohon', 'itu', 'berbuah', 'lebat'], ['tang', 'iku', 'berbuah', 'lebat']))
text.append(AlignedSent(['motor', 'itu', 'sangat', 'bagus'], ['motor', 'iku', 'sangat', 'bagus']))
text.append(AlignedSent(['harganya', 'sangat', 'mahal'], ['harganya', 'pisan', 'mahal']))
text.append(AlignedSent(['celana', 'itu', 'dari', 'bandung'], ['cali', 'iku', 'dari', 'bandung']))
text.append(AlignedSent(['hewan', 'ini', 'kurus'], ['sato', 'iki', 'kurus', 'begang']))
text.append(AlignedSent(['rumah', 'ini', 'besar'], ['imah', 'iki', 'gede']))
text.append(AlignedSent(['warung', 'ini', 'kecil'], ['warung', 'iki', 'cilik']))
text.append(AlignedSent(['penggaris', 'ini', 'pendek'], ['jidan', 'iku', 'pendek']))
text.append(AlignedSent(['sepatu', 'itu', 'sangat', 'bersih'], ['sap', 'iku', 'sangat', 'bersih']))
text.append(AlignedSent(['mobil', 'itu', 'berwarna', 'merah'], ['mot', 'iku', 'berwarna', 'merah']))
text.append(AlignedSent(['pintu', 'itu', 'tertutup'], ['panto', 'iku', 'tertutup']))
text.append(AlignedSent(['baju', 'itu', 'kusut'], ['baju', 'iku', 'karijut']))
text.append(AlignedSent(['kaca', 'itu', 'sangat', 'tebal'], ['kaca', 'iku', 'pisan', 'kaca']))
```

## 2. skor probability alignment 3 pasang kata.

The screenshot displays a Jupyter Notebook environment with a file named `alignment.ipynb`. The notebook contains a series of code cells and markdown cells. The code cells define an `IBMModel1` and use its `translation_table` to calculate probability alignments for three word pairs: "mobil" vs "mobil", "rumah" vs "omah", and "buku" vs "buku". The output cells show the resulting scores: 0.324, 0.465, and 0.482 respectively.

```
157 #%%
158 com_ibm1 = IBMModel1(text, 5)
159
160 #%% md
161
162 Coba cek skor probability alignment 2 kata
163
164 #%%
165
166 print(round(com_ibm1.translation_table["mobil"]["mobil"], 3) )
167
168 #%% md
169
170 Coba cek skor probability alignment 2 kata
171
172 #%%
173
174 print(round(com_ibm1.translation_table["rumah"]["omah"], 3) )
175
176 #%%
177
178 print(round(com_ibm1.translation_table["buku"]["buku"], 3) )
179
180 #%%
```

Coba lakukan pelatihan dari korpus yang didefinisikan sendiri

```
77 com_ibm1 = IBMModel1(text, 5)
```

Coba cek skor probability alignment 2 kata

```
78 print(round(com_ibm1.translation_table["mobil"]["mobil"], 3) )
0.324
```

Coba cek skor probability alignment 2 kata

```
79 print(round(com_ibm1.translation_table["rumah"]["omah"], 3) )
0.465
```

```
80 print(round(com_ibm1.translation_table["buku"]["buku"], 3) )
0.482
```

Jupyter Server started at <http://localhost:8888> // Open in Browser (24 minutes ago)

169.7 LF UTF-8 4 spaces Python 3.7

20:38 04/12/2020