Московский государственный технический университет им. Н.Э. Баумана
Кафедра «Системы обработки информации и управления»

Лабораторная работа №5
по дисциплине
«Методы машинного обучения»

Выполнил:
студент группы ИУ5И-21М
Мьоу Зо У

Москва — 2021 г.

1. **Цель лабораторной работы:**изучение методов предобработки текстов.

Задание:

Для произвольного предложения или текста решите следующие задачи:

1. Токенизация.
2. Частеречная разметка.
3. Лемматизация.
4. Выделение (распознавание) именованных сущностей.
5. Разбор предложения.

Текст программы и экранные формы

```
In [1]: !pip install -U pip setuptools wheel
        !pip install -U spacy
        !python -m spacy download en_core_web_sm
```

```
Collecting pip
  Downloading https://files.pythonhosted.org/packages/cd/6f/43037c7bcc8bd8ba7c9074256b1a11596daa15555808ec748048c1507f08/pip-21.1.1-py3-n
one-any.whl (1.5MB)
     |████████████████████████████████| 1.6MB 6.6MB/s
Collecting setuptools
  Downloading https://files.pythonhosted.org/packages/d0/15/5041473f5d142ee93bf1593deb8f932e27a078f6f04e2020cf44044f72c5/setuptools-56.2.
0-py3-none-any.whl (785kB)
     |████████████████████████████████| 788kB 31.4MB/s
Requirement already up-to-date: wheel in /usr/local/lib/python3.7/dist-packages (0.36.2)
ERROR: datascience 0.10.6 has requirement folium==0.2.1, but you'll have folium 0.8.3 which is incompatible.
Installing collected packages: pip, setuptools
  Found existing installation: pip 19.3.1
    Uninstalling pip-19.3.1:
      Successfully uninstalled pip-19.3.1
  Found existing installation: setuptools 56.1.0
    Uninstalling setuptools-56.1.0:
      Successfully uninstalled setuptools-56.1.0
Successfully installed pip-21.1.1 setuptools-56.2.0
Requirement already satisfied: spacy in /usr/local/lib/python3.7/dist-packages (2.2.4)
Collecting spacy
  Downloading spacy-3.0.6-cp37-cp37m-manylinux2014_x86_64.whl (12.8 MB)
     |████████████████████████████████| 12.8 MB 80 kB/s
Collecting typer<0.4.0,>=0.3.0
  Downloading typer-0.3.2-py3-none-any.whl (21 kB)
Requirement already satisfied: blis<0.8.0,>=0.4.0 in /usr/local/lib/python3.7/dist-packages (from spacy) (0.4.1)
Collecting thinc<8.1.0,>=8.0.3
  Downloading thinc-8.0.3-cp37-cp37m-manylinux2014_x86_64.whl (1.1 MB)
     |████████████████████████████████| 1.1 MB 38.0 MB/s
Collecting spacy-legacy<3.1.0,>=3.0.4
  Downloading spacy_legacy-3.0.5-py2.py3-none-any.whl (12 kB)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.7/dist-packages (from spacy) (20.9)
Requirement already satisfied: numpy>=1.15.0 in /usr/local/lib/python3.7/dist-packages (from spacy) (1.19.5)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.7/dist-packages (from spacy) (2.0.5)
Requirement already satisfied: tqdm<5.0.0,>=4.38.0 in /usr/local/lib/python3.7/dist-packages (from spacy) (4.41.1)
```

In [3]:
```
from spacy.lang.en import English
import spacy
```

In [4]:
```
text1 ='Bird Talker is distinguished by intelligence and ingenuity'
```

In [6]:
```
nlp = spacy.load('en_core_web_sm')
spacy_text1 = nlp(text1)
spacy_text1
```

Out[6]:
```
Bird Talker is distinguished by intelligence and ingenuity
```

In [7]:
```
for t in spacy_text1:
    print(t)
```

```
Bird
Talker
is
distinguished
by
intelligence
and
ingenuity
```

In [8]:
```
for token in spacy_text1:
    print('{} - {} - {}'.format(token.text, token.pos_, token.dep_))
```

```
Bird - PROPN - compound
Talker - PROPN - nsubjpass
is - AUX - auxpass
distinguished - VERB - ROOT
by - ADP - agent
intelligence - NOUN - pobj
and - CCONJ - cc
ingenuity - NOUN - conj
```

In [9]:
```
for token in spacy_text1:
      print(token, token.lemma, token.lemma_)
```

```
Bird 15457739884780427154 Bird
Talker 17522498178439073006 Talker
is 10382539506755952630 be
distinguished 1025273512777152759 distinguish
by 16764210730586636600 by
intelligence 11044490816763727375 intelligence
and 2283656566040971221 and
ingenuity 6309617539516521249 ingenuity
```

In [10]:
```
for ent in spacy_text1.ents:
    print(ent.text, ent.label_)
```

```
Bird Talker PERSON
```

In [11]:
```
from spacy import displacy
displacy.render(spacy_text1, style='ent', jupyter=True)
```

is distinguished by intelligence and ingenuity

In [12]:
```
from spacy import displacy
```

In [13]:
```
displacy.render(spacy_text1, style='dep', jupyter=True)
```