

3. Rozpoznaj jednostki nazewnicze w dokumentach, korzystając z API Clarin oraz wykorzystując model n82.

```
1. lpmn="any2txt|wcrft2|liner2({\"model\":\"n82\"}) "
```

7. Przedstaw 100 najczęściej rozpoznanych wyrażen, wraz z podaniem liczby ich wystąpień oraz kategorii semantycznej (tzn. niskopoziomowej (drobnoziarnistej) klasy przypisanej przez NER).

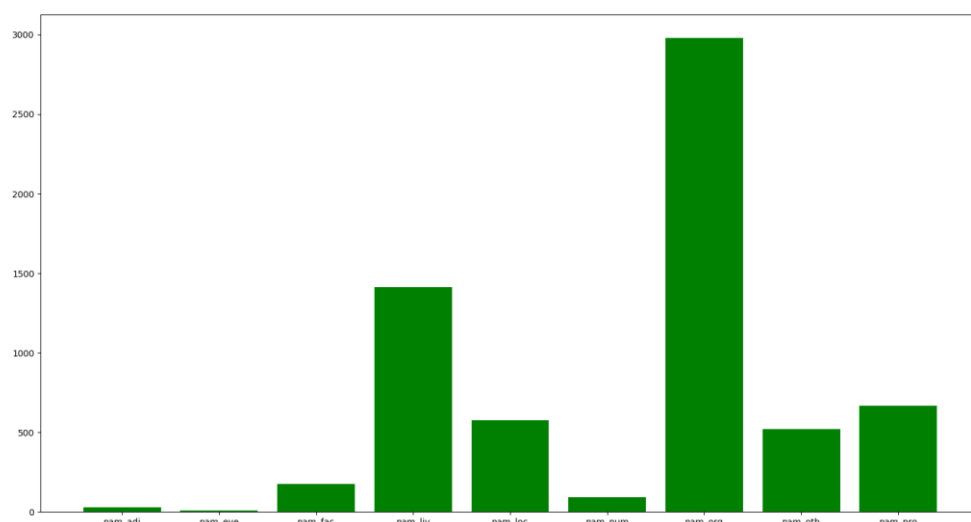
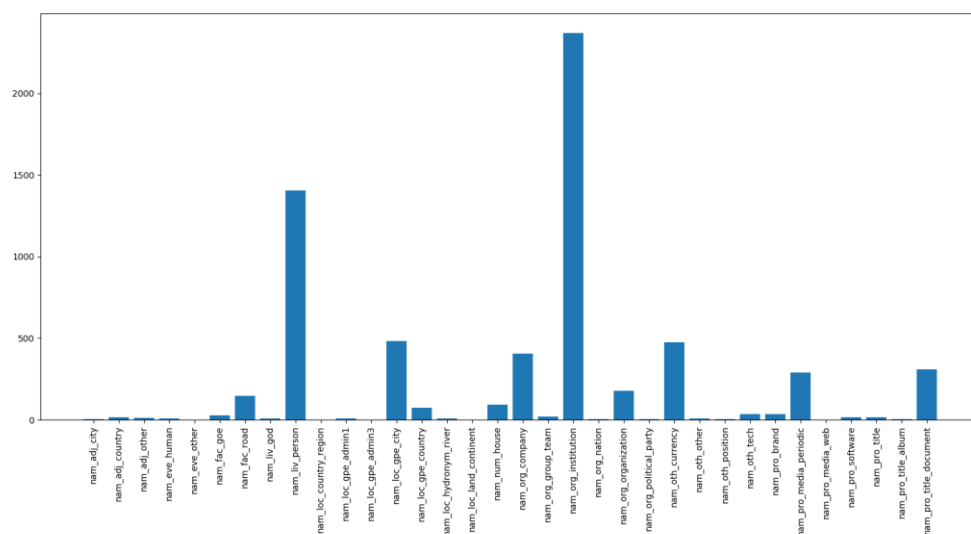
((('nam\_oth\_currency', 'zł'), 453)  
((('nam\_pro\_media\_periodic', 'Dz . U .'), 250)  
((('nam\_org\_institution', 'Sąd Najwyższy'), 236)  
((('nam\_org\_institution', 'Sądu Najwyższego'), 199)  
((('nam\_org\_institution', 'Sąd Okręgowy'), 174)  
((('nam\_liv\_person', 'Zamawiającego'), 164)  
((('nam\_org\_institution', 'Sąd Apelacyjny'), 126)  
((('nam\_org\_institution', 'Sądu Okręgowego'), 112)  
((('nam\_org\_institution', 'Sądu Apelacyjnego'), 85)  
((('nam\_org\_institution', 'Urzędu Zamówień Publicznych'), 82)  
((('nam\_pro\_title\_document', 'Prawo zamówień publicznych'), 77)  
((('nam\_liv\_person', 'Najwyższy'), 75)  
((('nam\_liv\_person', 'Zamawiający'), 74)  
((('nam\_loc\_gpe\_city', 'Warszawa'), 73)  
((('nam\_loc\_gpe\_city', 'Warszawie'), 66)  
((('nam\_org\_institution', 'Trybunał Konstytucyjny'), 57)  
((('nam\_org\_institution', 'Skarb Państwa'), 56)  
((('nam\_org\_institution', 'Sąd Rejonowy'), 50)  
((('nam\_org\_institution', 'Skarbu Państwa'), 50)  
((('nam\_org\_institution', 'Trybunału Konstytucyjnego'), 48)  
((('nam\_org\_institution', 'Prezesa Urzędu Zamówień Publicznych'), 43)  
((('nam\_loc\_gpe\_country', 'Polsce'), 43)  
((('nam\_org\_institution', 'TK'), 43)  
((('nam\_org\_institution', 'Sądu Rejonowego'), 42)  
((('nam\_liv\_person', 'Najwyższego'), 40)  
((('nam\_org\_institution', 'Izbie Cywilnej'), 35)  
((('nam\_org\_institution', 'Krajowa Izba Odwoławcza'), 34)  
((('nam\_liv\_person', 'Odwołującego'), 34)  
((('nam\_liv\_person', 'Scott Wilson Ltd'), 34)  
((('nam\_org\_institution', 'ETPC'), 33)  
((('nam\_org\_institution', 'IMIENIU RZECZYPOSPOLITEJ POLSKIEJ Dnia'), 31)  
((('nam\_liv\_person', 'Al'), 28)  
((('nam\_pro\_title\_document', 'Kodeks postępowania cywilnego'), 27)  
((('nam\_liv\_person', 'Odwołujący'), 25)

((('nam\_loc\_gpe\_city', 'Kraków'), 23)  
((('nam\_org\_organization', 'C'), 21)  
((('nam\_liv\_person', 'Zamawiającemu'), 20)  
((('nam\_liv\_person', 'Jana Pawła II'), 19)  
((('nam\_pro\_title\_document', 'Kodeks cywilny'), 18)  
((('nam\_oth\_currency', 'złotych'), 17)  
((('nam\_org\_institution', 'Izbie Pracy'), 16)  
((('nam\_org\_institution', 'Ubezpieczeń Społecznych i Spraw Publicznych'), 16)  
((('nam\_org\_institution', 'Sądowi Apelacyjnemu'), 16)  
((('nam\_liv\_person', 'Wykonawca'), 15)  
((('nam\_liv\_person', 'Skarżący'), 15)  
((('nam\_loc\_gpe\_city', 'Krakowie'), 14)  
((('nam\_liv\_person', 'Spółki'), 14)  
((('nam\_org\_institution', 'Sądowi Okręgowemu'), 14)  
((('nam\_pro\_title\_document', 'Kodeksu cywilnego'), 14)  
((('nam\_org\_company', 'CONTROL PROCESS'), 14)  
((('nam\_loc\_gpe\_city', 'Izby'), 13)  
((('nam\_liv\_person', 'M'), 13)  
((('nam\_org\_institution', 'Sąd'), 13)  
((('nam\_liv\_person', 'Gerard Bieniek'), 13)  
((('nam\_loc\_gpe\_city', 'Szczecin'), 12)  
((('nam\_loc\_gpe\_city', 'Rzeszowie'), 12)  
((('nam\_org\_institution', 'Zakład Pracy Chronionej'), 12)  
((('nam\_liv\_person', 'L . Wilk'), 12)  
((('nam\_loc\_gpe\_city', 'Warszawy'), 12)  
((('nam\_org\_institution', 'Sądu Okręgowego - Sądu Pracy i Ubezpieczeń Społecznych'), 11)  
((('nam\_org\_institution', 'Zakładu Ubezpieczeń Społecznych'), 11)  
((('nam\_pro\_title\_document', 'Konstytucji Rzeczypospolitej Polskiej'), 11)  
((('nam\_num\_house', '17'), 11)  
((('nam\_org\_company', 'Microsoft'), 11)  
((('nam\_loc\_gpe\_city', 'Tarnów'), 11)  
((('nam\_liv\_person', 'Edward Solak'), 11)  
((('nam\_org\_institution', 'Domu Pomocy Społecznej'), 10)  
((('nam\_org\_institution', 'Funduszu Ubezpieczeń Społecznych'), 10)  
((('nam\_liv\_person', 'Krzysztof Strzelczyk'), 10)  
((('nam\_loc\_gpe\_city', 'Katowicach'), 10)  
((('nam\_pro\_media\_periodic', 'CONSTAL'), 10)  
((('nam\_liv\_person', 'Andrzej Konieczka'), 10)  
((('nam\_loc\_gpe\_city', 'Zblewo'), 10)  
((('nam\_fac\_road', 'Kościerska'), 10)  
((('nam\_org\_institution', 'C'), 9)  
((('nam\_oth\_tech', 'C'), 9)  
((('nam\_org\_organization', 'K'), 9)  
((('nam\_loc\_gpe\_city', 'Białymstoku'), 9)  
((('nam\_loc\_gpe\_city', 'Wrocław'), 9)

(('nam\_liv\_person', 'J . Zagrodnik'), 9)  
 (('nam\_liv\_person', 'A . Światłowski'), 9)  
 (('nam\_pro\_title\_document', 'Konwencji'), 9)  
 (('nam\_org\_institution', 'Ministra Odbudowy'), 9)  
 (('nam\_loc\_gpe\_city', 'Tarnowie'), 9)  
 (('nam\_liv\_person', 'Magdalena Pazura'), 8)  
 (('nam\_org\_company', 'ComArch S . A .'), 8)  
 (('nam\_loc\_gpe\_city', 'Szczecinie'), 8)  
 (('nam\_pro\_brand', 'Netia S'), 8)  
 (('nam\_org\_institution', 'Sąd Okręgowy - Sąd Pracy i Ubezpieczeń Społecznych'), 8)  
 (('nam\_pro\_title\_document', 'rozporządzenia Rady Ministrów'), 8)  
 (('nam\_org\_institution', 'Sąd Okręgowy w K'), 8)  
 (('nam\_pro\_title\_document', 'ustawy o systemie ubezpieczeń społecznych'), 8)  
 (('nam\_org\_institution', 'Sądu Okręgowego w K'), 8)  
 (('nam\_fac\_road', 'Królowej Korony Polskiej'), 8)  
 (('nam\_org\_company', 'SCETAUROUTE Spółka Akcyjna Oddział'), 8)  
 (('nam\_loc\_hydronym\_river', 'Inwestycje'), 8)  
 (('nam\_num\_house', '18'), 8)  
 (('nam\_liv\_person', 'Piotra Ś'), 8)  
 (('nam\_liv\_person', 'Bogdan'), 8)  
 (('nam\_org\_organization', 'OTK ZU nr 1'), 8)

6. Przedstaw licznosc rozpoznanych klas w postaci dwóch wykresów:

Counter({'nam\_org\_institution': 2370, 'nam\_liv\_person': 1405, 'nam\_loc\_gpe\_city': 483,  
 'nam\_oth\_currency': 475, 'nam\_org\_company': 404, 'nam\_pro\_title\_document': 308,  
 'nam\_pro\_media\_periodic': 288, 'nam\_org\_organization': 179, 'nam\_fac\_road': 148,  
 'nam\_num\_house': 92, 'nam\_loc\_gpe\_country': 73, 'nam\_oth\_tech': 35, 'nam\_pro\_brand': 35,  
 'nam\_fac\_goe': 27, 'nam\_org\_group\_team': 19, 'nam\_pro\_title': 17, 'nam\_adj\_country': 14,  
 'nam\_pro\_software': 14, 'nam\_adj\_other': 11, 'nam\_loc\_hydronym\_river': 9, 'nam\_liv\_god': 9,  
 'nam\_oth\_other': 7, 'nam\_loc\_gpe\_admin1': 7, 'nam\_eve\_human': 7, 'nam\_adj\_city': 5,  
 'nam\_org\_nation': 4, 'nam\_oth\_position': 3, 'nam\_pro\_title\_album': 3, 'nam\_org\_political\_party': 2,  
 'nam\_eve\_other': 1, 'nam\_loc\_country\_region': 1, 'nam\_loc\_gpe\_admin3': 1, 'nam\_pro\_media\_web':  
 1, 'nam\_loc\_land\_continent': 1})



8. Przedstaw 10 najczęstszych wyrażeń, dla każdej wysokopoziomowej (zgrubnej) klasy wyrażeń.

nam\_org

('Sąd Najwyższy', 236)

('Sądu Najwyższego', 199)

('Sąd Okręgowy', 174)

('Sąd Apelacyjny', 126)

('Sądu Okręgowego', 112)

('Sądu Apelacyjnego', 85)

('Urzędu Zamówień Publicznych', 82)

('Trybunał Konstytucyjny', 57)

('Skarb Państwa', 56)

('Sąd Rejonowy', 50)

---

nam\_liv

('Zamawiającego', 164)

('Najwyższy', 75)

('Zamawiający', 74)

('Najwyższego', 40)

('Odwołującego', 34)

('Scott Wilson Ltd', 34)

('Al', 28)

('Odwołujący', 25)

('Zamawiającemu', 21)

('Jana Pawła II', 19)

---

nam\_loc

('Warszawa', 73)

('Warszawie', 66)

('Polsce', 43)

('Kraków', 23)

('Krakowie', 14)

('Izby', 13)

('Szczecin', 12)

('Rzeszowie', 12)

('Warszawy', 12)

('Tarnów', 11)

---

nam\_fac

('Kościerska', 10)

('Królowej Korony Polskiej', 8)

('Puławska', 7)

('Wały Chrobrego', 6)

('Rejtana', 6)

('Grójecka', 5)

('Stawki', 5)

('Mickiewicza', 5)

('Kondratowicza', 5)

('Modlińska', 5)

---

nam\_num

('17', 11)

('18', 8)

('29', 7)

('21', 6)

('182', 5)

('6', 4)  
('128', 4)  
('2', 4)  
('1', 4)  
('8', 4)

---

nam\_oth  
('zł', 453)  
('złotych', 17)  
('C', 10)  
('VAT', 8)  
('Internet', 6)  
('PN', 3)  
('PLN', 3)  
('euro', 2)  
('L', 2)  
('RP', 2)

---

nam\_pro  
('Dz . U .', 250)  
('Prawo zamówień publicznych', 77)  
('Kodeks postępowania cywilnego', 27)  
('Kodeks cywilny', 18)  
('Kodeksu cywilnego', 14)  
('Konstytucji Rzeczypospolitej Polskiej', 11)  
('CONSTAL', 10)  
('Konwencji', 9)  
('Netia S', 8)  
('rozporządzenia Rady Ministrów', 8)

---

nam\_adj  
('polskiej', 4)  
('internetowej', 3)  
('europejskiej', 3)  
('warszawskiego', 3)  
('libijskiego', 2)  
('unijnych', 1)  
('libijski', 1)  
('libijskimi', 1)  
('libijska', 1)  
('amerykańskich', 1)

---

nam\_eve  
('wojny babka', 1)  
('OS - 1', 1)

('Kodeksu Karnego', 1)  
('Spółki', 1)  
('NTT Business W 9001', 1)  
('C 2', 1)  
('Świątami Wielkanocnymi', 1)  
('Normosan przeciwko Normovit', 1)

---

## 9. Oceń przydatność NERa dla danych jakimi są orzeczenia sądów.

W wynikach pojawiają się wyrażenia mocno związane z sądami. Po ekstrakcji informacji można wywnioskować jakiego typu jest to tekst.