



# NLP modellek – újszerű alkalmazások

Lukács András

lukacs@cs.elte.hu

ELTE Matematikai Intézet  
MI Kutatócsoport

MILab NLP megbeszélés – 2020. 12. 11.



# Alkalmazási területek - adatok

- Idősorok
- IT logok
  - Biztonsági logok
    - Biztonsági események felismerése, jellemzése
  - Működési logok
    - Zavarok felismerése és javítása
    - Működés optimalizálás



# Alkalmazási területek - adatok

- Biológiai szekvenciák
  - DNS/RNS/fehérje szekvenciák
    - Klasszifikációs feladatok
    - Motívumok keresése
      - Gyakori, konzervatív részletek – funkciót jelent
    - Generálás?
  - Kismolekulák
    - SMILES és hasonló szekvenciális leírások
    - Generálás?



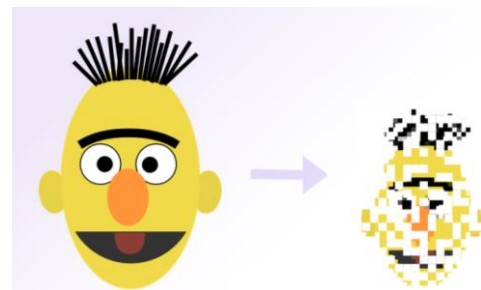
# Alkalmazási területek - adatok

- Szoftver forráskódok és kommentek
  - Kód → leírás
    - Kulcsszavazás
    - Automatikus kommentezés
  - Leírás → Kód
    - Kódbázisban keresés
    - Kódgenerálás



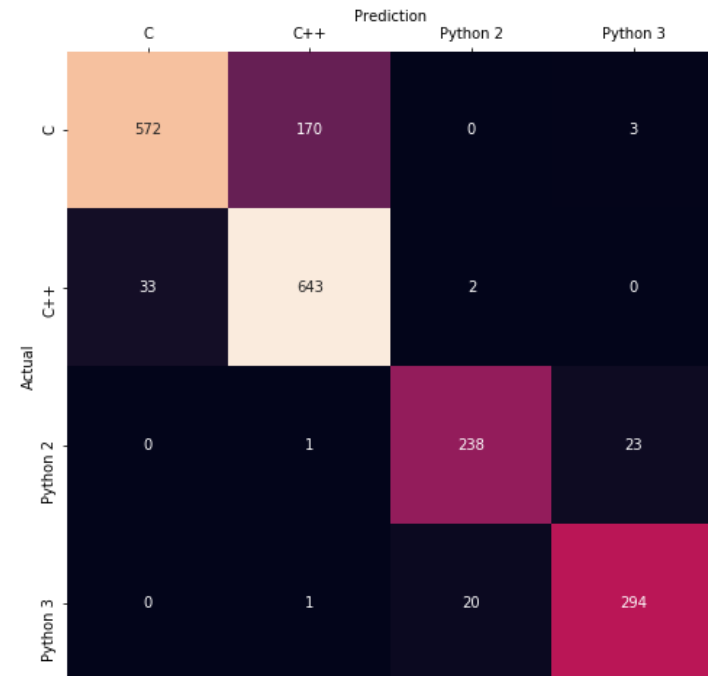
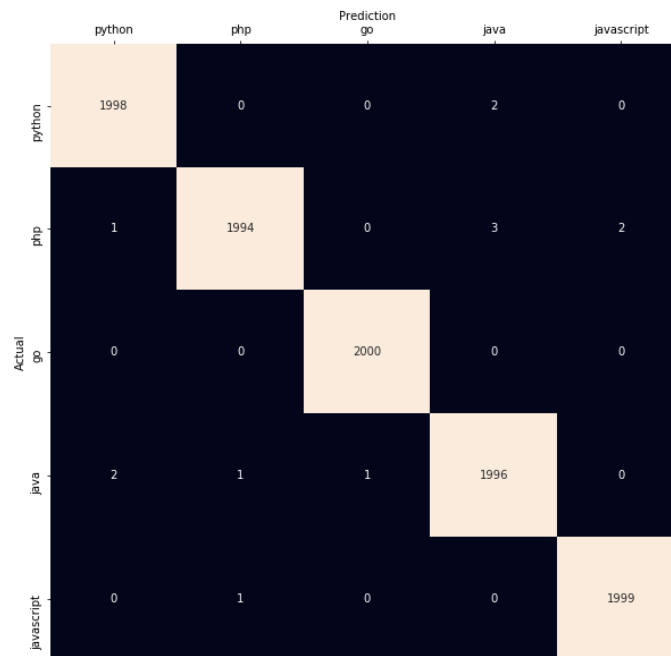
# NLP modellek

- Rekurrens hálózatok
  - LSTM, GRU
- Konvolúciós megközelítések
  - 1D CNN
    - Egyedi megoldások
    - Automatikus optimalizálás
  - Gráfkonvolúciók
- Transzformerek
  - BERT és társaik
    - Domain specifikus modellek
    - Kisebb, tömörebb modellek



# Programnyelvek felismerése

- CodeSearchNet és codeforces corpus
- LSTM



# Fehérjék osztályozása

- Pfam (generált osztályok HMM-mel)
- 1D CNN

