

## Mesterséges intelligencia és gépi tanulás tanulmányi verseny MIGT-2020

### “Egy mesterséges intelligencia könyv rekonstrukciója”

A verseny során mesterséges intelligenciához kötődő magyar nyelvű szövegtöredékeket kell rendszerezni Stuart Russell és Peter Norvig által írt Mesterséges intelligencia - modern megközelítésben című könyvének második kiadása alapján. A szövegtöredékek egy része magából a könyvből származik, más része csupán a mesterséges intelligencia tárgyköréhez kapcsolódik. A feladatban a szövegtöredékek egy részéhez: (1) a fejezetének száma, (2) a fejezetbeli pozíciója és (3) a könyvben való előfordulásának ténye elérhető. A cél egy olyan mesterséges intelligencia és gépi tanulás alapuló megoldás létrehozása, amely csak ezekre a már ismert rendszerezettséggű szövegtöredékekre támaszkodva lehetővé teszi további szövegtöredékek ilyen hármas szisztémában való rendszerezését. **A megoldásban a könyv elektronikusan elérhető verzióinak használata semmilyen formában nem megengedett!**

A megoldások kiértékelésében a könyvben való előfordulás bináris predikciója és a könyvbeli pozíciók hierarchikus predikciója egyenlő súllyal számít (részleteket lásd Teljesítménymetrikáknál lent).

#### Bemeneti információk

Az adatokat egy **UTF-8** kódolású szöveges file tartalmazza, tab-szeperált formában. A file első sora az oszlopok címkeit tartalmazza, ezután minden további sora egy tanító mintát tartalmaz, amely az alábbi oszlopokból áll:

1. **chapter#** - A bekezdés az MI könyv melyik fejezetéből származik. [1-27] közötti egész szám.
2. **sectionfloat** - A bekezdés MI könyv fejezetén belül sorrendben hol található. [0-1] közötti valós szám
3. **isextra** - A bekezdés a nyomtatott MI könyvből (0), vagy a hozzá készített segédanyagokból származik (1). Bináris, 0 vagy 1.
4. **paragraph** - A könyv egy magyar nyelvű szövegtöredéke. Tab, újsor karakterek nem szerepelnek benne, legalább 200 karakter hosszú string, legfeljebb 10000 karakter hosszú.

A file elérhetősége: [http://phoenix.mit.bme.hu:49080/kgt/MIGT\\_2020\\_traindata\\_V1.tsv](http://phoenix.mit.bme.hu:49080/kgt/MIGT_2020_traindata_V1.tsv)

#### Az elvárt kimenet

Az elvárt kimeneti predikció ez a három “rendszerezési” jellemző egy 20 órakor kiadott VALIDATION adathalmazon, nevezetesen a következő megnevezésekkel:

- a skalár CHAPTER,
- a skalár sectionfloat,
- A bináris isExtra

A validációs adathalmaz elérhetősége:

[http://phoenix.mit.bme.hu:49080/kgt/MIGT\\_2020\\_validation\\_V1.tsv](http://phoenix.mit.bme.hu:49080/kgt/MIGT_2020_validation_V1.tsv)

#### Az elvárt kimeneti formátum

A predikciókat szintaktikailag egy RESULTS\_versenyzoineiNEPTUNKod.TSV file-ban, (tab szeparált) várjuk, amelyben rendre a prediktált a CHAPTER értéket, a POSITION értéket és az AIMA bináris (0/1) értéket, majd utolsóként az eredeti szövegtöredéket.

### **Teljesítménymetrikák (a RESULTS\*.TSV alapján)**

- Bináris félreosztályozási hiba az AIMA alapján.
- Négyzetes hiba a CHAPTER alapján.
- Négyzetes hiba a POSITION alapján.

A versenyen beadott megoldásokat a három teljesítménymetrika alapján külön-külön sorrendezzük, döntetlent megengedve, majd az egyes megoldások összesített helyezését az elért három sorrendjének az átlaga határozza meg.

### **Beadás**

A MIT tanszék HF kezelői rendszerén (<https://hf.mit.bme.hu/>) kérjük feltölteni a következőket egy MIGT\_versenyzoineiNEPTUNKod gz/zip file-ban:

- A validációs halmazra adott predikciókat a RESULTS\*TSV file-ban.
- A megoldást adó forráskódot tömörítve.
- A megoldás elméleti háttérének összefoglalóját 50-500 karakterben szóközökkel együtt.
- A megoldás lefuttatásához szükséges technikai feljegyzéseket 50-500 karakterben.

A ZIP file-t a MIT tanszék HF kezelői rendszerén keresztül (<https://hf.mit.bme.hu/>) kérjük feltölteni "MIGT2020 tanulmányi verseny" címszó alatt 21 óráig. **A 21 óra után feltöltött feladatokat érvénytelennek tekintjük!**

### **A bemeneti adatok és változóinak a részletes leírásai, és váz a versenyfeladathoz**

A Google Colab rendszeren belül egy Python3 környezetet használhattok, ahol könnyen elkészíthetitek és dokumentálhatjátok a megoldásaitokat. Néhány hasznos függvényt is előkészítettünk benne. A használatához készítsetek másolatot az alábbi vázból:

[https://colab.research.google.com/drive/1vM9AUa4D83iGmaPVK2r7CJj\\_nkngQ3lr#scrollTo=y1S9cPzX3mXH](https://colab.research.google.com/drive/1vM9AUa4D83iGmaPVK2r7CJj_nkngQ3lr#scrollTo=y1S9cPzX3mXH)

Jó versenyzést kívánunk!

**ComBine.lab (Computational Biomedicine workgroup)**  
**Artificial Intelligence group**  
**Department of Measurement and Information Systems**