

Jegyzet a 7. óra anyagához

Simon Eszter

2020. március 30.

Ezen az órán a gépi tanulásról lesz szó. Ez a jegyzet a sorvezető az órához. Pirossal fogom jelölni a jegyzet adott pontján elvégzendő feladatot. A feladatokat annyi, hogy ezt a jegyzetet olvassátok, és a jelölt feladatokat sorban elvégezzétek.

A tulajdonnév-felismerés (*named entity recognition*, *NER*) lesz az a téma, ami az óra anyagán végigvezet minket; a példák is mindig erről a területről jönnek. Ez az órai anyag nagyjából a disszertációm 5. fejezetének kivonatolt változata. A disszertációm elérhető a <https://repozitorium.omikk.bme.hu/bitstream/handle/10890/1250/ertekezes.pdf> URL-en keresztül.

Olvassátok el az 5. fejezetet!

Ezután nyissátok meg a 7. órához tartozó szlájdkokat. **Nézzétek át a szlájdkokat a 17. szlájdig!**

Magyarázat a következő fejezet címéhez: a HunTag egy szekvenciális tagger, ami többféle szekvenciális címkézési feladatra használható, mint amilyen például a tulajdonnév-felismerés is. Az eszköz legújabb verziója HunTag3 névre hallgat, és elérhető a <https://github.com/dlt-rilmta/HunTag3> GitHub repóban.

Nézzétek meg a readme-jét, majd térjete tovább a szlájdkokhoz!

A 22. és 23. szlájdkokon található példa HMM és a Viterbi-algoritmus megértéséhez **olvassátok el** Kovásznai Gergely: Párbeszédes rendszerek című könyvének 3. fejezetének 1.1. alfejezetét, amely a Digitális Tankönyvtárból érhető el: <https://dtk.tankonyvtar.hu/xmlui/handle/123456789/8390>!

Olvassátok tovább a szlájdkokat!

Megjegyzés a 28. szlájdkhoz: „hagyományos” gépi tanulás alatt a neurális forradalom előtti gépi tanulási rendszereket értem. A neurális gépi tanulásról majd későbbi órákon lesz szó.

Házi feladat: Milyen kedvem van az egyes napokon, ha mindennap többórás házi feladatot adok? A példa HMM-nél megadott értékek és számítási módszer felhasználásával egyrészt kérek egy állapotsekvenciát a hét öt napjára, másrészt a teljes Trellis-diagramot a számokkal. Határidő: április 20.