

A TANTÁRGY ADATLAPJA

(Természetesnyelv-feldolgozási módszerek)

Egyetemi tanév: 2025–2026

1. A képzési program adatai

1.1. Felsőoktatási intézmény	Babeş-Bolyai Tudományegyetem
1.2. Kar	Matematika és Informatika Kar
1.3. Intézet	Magyar Matematika és Informatika Intézet
1.4. Szakterület	Informatika
1.5. Képzési szint	Mestéri képzés
1.6. Tanulmányi program / Képesítés	Adatelemzés és modellezés / Analiza datelor și modelare Data analysis and modelling
1.7. Képzési forma	Nappali

2. A tantárgy adatai

2.1. A tantárgy neve		Természetesnyelv-feldolgozási módszerek / Metode de prelucrare a limbajului natural / Methods of natural language processing				A tantárgy kódja	MMM8081
2.2. Az előadásért felelős tanár neve			dr. Bodó Zalán-Péter egyetemi docens				
2.3. A szemináriumért felelős tanár neve			dr. Bodó Zalán-Péter egyetemi docens				
2.4. Tanulmányi év	1	2.5. Félév	2	2.6. Értékelés módja	V	2.7. Tantárgy típusa	kötelező- alaptárgy

3. Teljes becsült idő (az oktatási tevékenység féléves óraszama)

3.1. Heti óraszám	5	melyből: 3.2 előadás	2	3.3 szeminárium/labór/projekt	1+2+0
3.4. Tantervben szereplő összóraszám	70	melyből: 3.5 előadás	28	3.6 szeminárium/labór	42
Az egyéni tanulmányi idő (ET) és az önképzési tevékenységekre (ÖT) szánt idő elosztása:					óra
A tankönyv, a jegyzet, a szakirodalom vagy saját jegyzetek tanulmányozása					50
Könyvtárban, elektronikus adatbázisokban vagy terepen való további tájékozódás					30
Szemináriumok / laborok, házi feladatok, portfóliók, referátumok, esszék kidolgozása					40
Egyéni készségfejlesztés (tutorálás)					6
Vizsgák					4
Más tevékenységek:					0
3.7. Egyéni tanulmányi idő (ET) és önképzési tevékenységekre (ÖT) szánt idő összóraszama			130		
3.8. A félév összóraszama			200		
3.9. Kreditszám			8		

4. Előfeltételek (ha vannak)

4.1. Tantervi	Nincs
4.2. Kompetenciabeli	Algoritmika, programozási készségek, matematikai alapismeretek (algebra, valószínűségszámítás).

5. Feltételek (ha vannak)

5.1. Az előadás lebonyolításának feltételei	Az előadásokhoz videoprojektor
---	--

	szükségeltetik. • A példák kifejtéséhez tábla szükséges.
5.2. A szeminárium / labor lebonyolításának feltételei	• A szemináriumok és laborok lebonyolításához az előadáséhoz hasonló feltételek szükségesek.

6. Elsajátítandó jellemző kompetenciák

Szakmai/kulcs-kompetenciák	- A számítógépes nyelvészeti fogalmak és algoritmusok ismerete. - Korpuszok használata és tervezése. - Algoritmusok elemzése és fejlesztése.
Transzverzális kompetenciák	- Önálló tanulás - Munkamódszerek, módszertani kompetenciák - Kritikus gondolkodás és reflexió

7. A tantárgy célkitűzései (az elsajátítandó jellemző kompetenciák alapján)

7.1 A tantárgy általános célkitűzése	A tantárgy célja a természetes nyelven írt (elektronikus) szövegek feldolgozási módszereinek ismertetése elméleti és gyakorlati szempontból.
7.2 A tantárgy sajátos célkitűzései	- A természetes nyelvfeldolgozás foglmainak és módszereinek ismerete és alkalmazása: - Számítógépes nyelvészeti korpuszok - Szó alapú és n-gram modellek - Szintaktikai elemzések - Markov modellek - Számítógépes nyelvészeti feladatok megoldása: szófaji egyértelműsítés, jelentés-egyértelműsítés, koreferenciák feloldása, információkinyerés, kivonatolás, érzelem-elemzés, gépi fordítás stb.

8. A tantárgy tartalma

8.1 Előadás	Didaktikai módszerek	Megjegyzések
1. Természetes és formális nyelvek, a természetes nyelvfeldolgozás története és alapfogalmai.	tanári magyarázat, rávezetés, munkáltatás	
2. Alapvető szövegfeldolgozási módszerek. A szerkesztési távolság.	tanári magyarázat, rávezetés, munkáltatás	
3.	tanári magyarázat, rávezetés, munkáltatás	

Szövegekategorizálás. A Naiv Bayes módszer.		
4. Kiértékelési módszerek.	tanári magyarázat, rávezetés, munkáltatás	
5. Entrópia és hibajavító kódok a gépi tanulásban és természetesnyelv-feldolgozásban.	tanári magyarázat, rávezetés, munkáltatás	
6. Nyelvi modellek, n-grammok.	tanári magyarázat, rávezetés, munkáltatás	
7. Álhírek.	tanári magyarázat, rávezetés, munkáltatás	
8. Érzelelemelemzés, véleménybányászat.	tanári magyarázat, rávezetés, munkáltatás	
9. Szavak, mondatok, dokumentumok reprezentációja a természetesnyelv-feldolgozásban és információ-visszakeresésben I.	tanári magyarázat, rávezetés, munkáltatás	
10. Szavak, mondatok, dokumentumok reprezentációja a természetesnyelv-feldolgozásban és információ-visszakeresésben II.	tanári magyarázat, rávezetés, munkáltatás	
11. Szintaktikai elemzések.	tanári magyarázat, rávezetés, munkáltatás	
12. Dokumentumok automatikus kivonatolása (summarization).	tanári magyarázat, rávezetés, munkáltatás	
13. Információkinyerés (information extraction).	tanári magyarázat, rávezetés, munkáltatás	
14. Összefoglalás, ismétlés.	tanári magyarázat, rávezetés, munkáltatás	

Könyvészet:

[1] JURAFSKY M., MARTIN J.H. Speech and Language Processing. 3rd ed. (draft), 2024 (<https://web.stanford.edu/~jurafsky/slp3/>).

[2] TĂTAR D. Inteligență artificială. Aplicații în prelucrarea limbajului natural. Editura Albastră, Cluj

Napoca, 2003. [3] MANNING C.D., SCHÜTZE H. Foundations of statistical language processing. MIT Press, Cambridge, 1999. [4] TIKK DOMONKOS. Szövegbányászat. Typotex, Budapest, 2007. [5] Andriy Burkov. The Hundred-Page Language Models Book: hands-on with PyTorch. True Positive Inc., 2025.		
8.2 Szeminárium / Labor	Didaktikai módszerek	Megjegyzések
A szemináriumokon nagyrészt cikkeket olvasunk a hallgatókkal, megbeszéljük a kötelező projekteket (ezek némileg eltérhetnek a lennebb található projektektől), az azokkal kapcsolatos problémákat, a hallgatók a kitűzött határidőre bemutatják a kötelező évközi projekteket, majd a félév második részében a csoportos projektek bemutatói (NLP alkalmazás + bemutató) kerülnek sorra ezeken az órákon.	kooperatív csoportmunka, hallgatói kiselőadás, vita, megbeszélés	
Feladatok:		
Csoportos projekt (2-3 személy a projekt komplexitásától függően)	csoportos munka	
1. Sentiment analysis: saját szótár készítése és annak alkalmazási szövegek érzelmi elemzésében	individuális munka	
2. Wikification: szöveg automatikus wikifikációja	individuális munka	
3. Szóbeágyazások: saját neurális modell építése és betanítása	individuális munka	
<i>Könyvészet:</i> [1]-[5]+ [6] JACKSON P., MOULINIER I. Natural Language Processing for Online Applications. Text Retrieval, Extraction and Categorization. John Benjamins Publishing Company, 2002. [7] ABNEY S. Semisupervised learning for computational linguistics. Chapman & Hall/CRC, 2008. [8] FELDMAN R., SANGER J. The Text Mining Handbook. Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press, 2007.		

9. Az episztemikus közösségek képviselői, a szakmai egyesületek és a szakterület reprezentatív munkáltatói elvárásainak összhangba hozása a tantárgy tartalmával.

- Az előadás vázát a Dan Jurafsky és Christopher Manning "Natural Language Processing" c. Coursera kurzus képezi https://web.stanford.edu/~jurafsky/NLPCourseraSlides.html - Jelentős átfedés van a jelen tantárgy tartalma és a "Speech and Language Processing", illetve a "Foundations of statistical language processing" c. könyvek tartalma között (lásd ezeket a bibliográfiában), melyeket az NLP alapműveinek tekintjük.
--

10. Értékelés

Tevékenység típusa	10.1 Értékelési kritériumok	10.2 Értékelési módszerek	10.3 Aránya a végső jegyben
10.4 Előadás	Vizsgafeladatok	Írásbeli vizsga	40%
10.5 Szeminárium / Labor	Az évközi tevékenység keretében létrejött (csoportos) projekt (alkalmazás és bemutató)	A projekt kiértékelése	30%
	Az évközi tevékenység	A projekt kiértékelése	30%

	keretében létrejött egyéni projektek (csak alkalmazás)		
10.6 A teljesítmény minimumkövetelményei			
Az elvárt minimális tudás: • Alapvető NLP fogalmak ismerete. • Alapvető NLP algoritmusok ismerete. • Egy NLP feladat önálló megoldása. Az átmenő jegy feltételei: • A pontok felének összeszedése minden kiértékeléskor kötelező.			

11. SDG ikonok (Fenntartható fejlődési célok/ Sustainable Development Goals)

	A fenntartható fejlődés általános ikonja						
							

Kitöltés időpontja:
2025.04.30

Előadás felelőse:
dr. Bodó Zalán-Péter

Szeminárium felelőse:
dr. Bodó Zalán-Péter

Az intézeti jóváhagyás dátuma:
2025.04.30

Intézetigazgató:
dr. András Szilárd