

A TANTÁRGY ADATLAPJA

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	Mesteri
1.6 Szak / Képesítés	Adatelemzés és modellezés / Analiza datelor și modelare Data analysis and modelling

2. A tantárgy adatai

2.1 A tantárgy neve	Természetes nyelvfeldolgozási módszerek / Metode de prelucrare a limbajului natural / Methods of natural language processing						
2.2 Az előadásért felelős tanár neve	Bodó Zalán-Péter						
2.3 A szemináriumért felelős tanár neve	Bodó Zalán-Péter						
2.4 Tanulmányi év	2	2.5 Félév	2	2.6. Értékelés módja	vizsga	2.7 Tantárgy típusa	kötelező – alap

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

3.1 Heti óraszám	5	melyből: 3.2 előadás	2	3.3 szeminárium/labor/praktika	1+2
3.4 Tantervben szereplő össz-óraszám	70	melyből: 3.5 előadás	28	3.6 szeminárium/labor	42
A tanulmányi idő elosztása:					Óra
A tankönyv, a jegyzet, a szakirodalom vagy saját jegyzetek tanulmányozása					36
Könyvtárban, elektronikus adatbázisokban vagy terepen való további tájékozódás					36
Szemináriumok / laborok, házi feladatok, portfóliók, referátumok, esszék kidolgozása					40
Egyéni készségfejlesztés (tutorálás)					12
Vizsgák					6
Más tevékenységek:					
3.7 Egyéni munka össz-óraszama					130
3.8 A félév össz-óraszama					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	<ul style="list-style-type: none"> • Az előadásokhoz video-projektor szükséges. • A példák kifejtéséhez és illusztráció számára tábla szükséges.
5.2 A szeminárium / labor	<ul style="list-style-type: none"> • A laboratóriumi órák alatt a diákok a számítógépet, az oktató a táblát

lebonyolításának feltételei	használja.
-----------------------------	------------

6. Elsajátítandó jellemző kompetenciák

Szakmai kompetenciák	<ul style="list-style-type: none"> • A számítógépes nyelvészeti fogalmak és algoritmusok ismerete. • Korpuszok használata és tervezése. • Algoritmusok elemzése és fejlesztése.
Transzverzális kompetenciák	<ul style="list-style-type: none"> • Önálló tanulás • Munkamódszerek, módszertani kompetenciák • Kritikus gondolkodás és reflexió

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

7.1 A tantárgy általános célkitűzése	<ul style="list-style-type: none"> • A tantárgy célja a természetes nyelven írt (elektronikus) szövegek feldolgozási módszereinek ismertetése elméleti és gyakorlati szempontból.
7.2 A tantárgy sajátos célkitűzései	<ul style="list-style-type: none"> • Az természetes nyelvfeldolgozás fogalmainak és módszereinek ismerete és alkalmazása: <ul style="list-style-type: none"> ○ 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.

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. Számítógépes nyelvészeti korpuszok.	tanári magyarázat, rávezetés, munkáltatás	
3. Szótövesítő (stemming) algoritmusok: a Porter-féle algoritmus.	tanári magyarázat, rávezetés, munkáltatás	
4. Kollokációk a számítógépes nyelvészetben.	tanári magyarázat, rávezetés, munkáltatás	
5. N-gramok és n-gram modellek.	tanári magyarázat, rávezetés, munkáltatás	
6. Szintaktikai elemzések: környezetfüggetlen és	tanári magyarázat, rávezetés,	

valószínűségi környezetfüggetlen (PCFG) grammatikák, valószínűségi elemzési módszerek.	munkáltatás	
7. Markov modellek: Markov láncok, rejtett Markov modellek (HMM), a Viterbi algoritmus.	tanári magyarázat, rávezetés, munkáltatás	
8. Markov modellek: a Forward és a Baum–Welch algoritmus. Alkalmazás szófaji egyértelműsítésre (Part-Of-Speech (POS) Tagging).	tanári magyarázat, rávezetés, munkáltatás	
9. Szavak jelentés-egyértelműsítése (word sense disambiguation, WSD).	tanári magyarázat, rávezetés, munkáltatás	
10. Koreferenciák feloldása.	tanári magyarázat, rávezetés, munkáltatás	
11. Információkinyerési módszerek (information extraction).	tanári magyarázat, rávezetés, munkáltatás	
12. Dokumentumok automatikus kivonatolása (summarization).	tanári magyarázat, rávezetés, munkáltatás	
13. Érzelem-elemzés (sentiment analysis) természetes nyelvű szövegekben.	tanári magyarázat, rávezetés, munkáltatás	
14. A gépi fordítás módszerei.	tanári magyarázat, rávezetés, munkáltatás	

Könyvészet

- [1] TÁTAR D. *Inteligență artificială. Aplicații în prelucrarea limbajului natural*. Editura Albastră, Cluj-Napoca, 2003.
- [2] MANNING C.D., SCHÜTZE H. *Foundations of statistical language processing*. MIT Press, Cambridge, 1999.
- [3] JURAFSKY M., MARTIN J.H. *Speech and Language Processing*. Prentice-Hall, 2000.
- [4] JACKSON P., MOULINIER I. *Natural Language Processing for Online Applications. Text Retrieval, Extraction and Categorization*. John Benjamins Publishing Company, 2002.
- [5] ABNEY S. *Semisupervised learning for computational linguistics*. Chapman & Hall/CRC, 2008.

8.2 Szeminárium / Labor	Didaktikai módszerek	Megjegyzések
1. Helyesírás-ellenőrző és javító programok.	munkáltatás, demonstráció, példák	
2. A WordNet adatbázis.	munkáltatás, demonstráció, példák	
3. Hasonlósági mértékek, szavak és dokumentumok klaszterezése.	munkáltatás, demonstráció, példák	
4. Szavak jelentés-egyértelműsítése (word sense disambiguation, WSD).	munkáltatás, demonstráció, példák	
5. Koreferenciák feloldása.	munkáltatás, demonstráció, példák	

6. Az IBM Watson-ban használt módszerek.	munkáltatás, demonstráció, példák	
7. Összefoglalás, feladatok bemutatása.		Szemináriumi bemutatók, melyeket a diákok tartanak a témáról és a tartalomról történt egyeztetés után.

Könyvészet

[1]–[5] +

[6] BERRY M.W. (ed.) *Survey of Text Mining. Clustering, Classification and Retrieval*. Springer, 2003.

[7] BILISOLY R. *Practical Text Mining With Perl*. Wiley, 2008.

[9] FELDMAN R., SANGER J. *The Text Mining Handbook. Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, 2007.

[10] http://researcher.ibm.com/researcher/view_project.php?id=2099

[11] <http://www.ibm.com/smarterplanet/us/en/ibmwatson/>

[12] <http://cs.nyu.edu/courses/Fall12/CSCI-GA.3033-001/index.html>

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

- Az előadás vázát a New York University-n tartott "Statistical Natural Language Processing" c. kurzus képezi (<http://cs.nyu.edu/courses/Fall12/CSCI-GA.3033-001/index.html>).

10. Értékelés

Tevékenység típusa	10.1 Értékelési kritériumok	10.2 Értékelési módszerek	10.3 Aránya a végső jegyben
10.4 Előadás	Írásbeli vizsga a félév végén	Írásbeli vizsga	60%
10.5 Labor	Programozási feladatok bemutatása és bemutatók	A megoldások pontozása	40%
10.6 A teljesítmény minimumkövetelményei			
Kötelező a pontok felének összeszedése minden kiértékeléskor (évközi kiértékelés (laborgyakorlatok, szemináriumi bemutatók), végső vizsga).			

Kitöltés dátuma

2016.04.19

Előadás felelőse

Bodó Zalán-Péter

Labor / praktika felelőse

Bodó Zalán-Péter

Az intézeti jóváhagyás dátuma

.....

Intézetigazgató

András Szilárd