



Tantárgy kód

BMETE91MX12

Tantárgy azonosító adatok

1.	A tárgy címe	Bevezetés a nyelvmérnökségbe 1 – Szószint eszközök									
2.	A tárgy angol címe	Introduction to Language Engineering 1 – Processing of Words									
3.	A tárgy rövid címe	Nyelvmérnökségl	Követelmény	2	+	0	+	0	v	Kredit	3
4.	Ajánlott/kötelező tanulmányi rend										
	vagy	Tantárgy kód 1	Rövid cím 1	Tantárgy kód 2	Rövid cím 2	Tantárgy kód 3	Rövid cím 3				
4.1											
4.2											
4.3											
5.	Kizáró tantárgyak										
6.	A tantárgy felelős tanszéke	Algebra Tanszék									
7.	A tantárgy felelős oktatója	Dr. Rónyai Lajos	beosztása	egyetemi tanár							

Akkreditációs adatok

8.	Akkreditációra benyújtás időpontja	2010.10.13.	Akkreditációs bizottsági döntés időpontja	2010.12.20.
----	------------------------------------	--------------------	---	--------------------

Megjegyzések

Csak az űrlap fehéren hagyott mezőbe írjunk és a mezők között a **tabulátor** billentyűvel haladjunk! Ha egy kitöltött mezőből tabulátor billentyűvel lépünk ki, több más mező értéke automatikusan megváltozhat. Egy adott mezőre lépve, az állapotsorban megjelenő rövid, ill. az F1 gomb megnyomásakor kapható hosszabb leírás ad segítséget a kitöltéshez. A *tantárgy kódot* a dékáni hivatal adja.

1-2. sorok: A tárgy címének (max. 60 karakter) legalább egy karakterben különböznie kell minden más, Neptunban regisztrált tárgy címétől.

3. sor: A rövid cím jellegzetes, legfeljebb 16 karakter hosszúságú rövidítés. A követelmény eladás+gyakorlat+labor formátumú, az utolsó mező a félév végi számonkérés típusa (v,f,a vagy s, részletes információ az F1 gombra). A kredit megadásánál ügyelni kell arra, hogy az alább részletezett, a tantárgy elvégzéséhez átlagosan szükséges tanulmányi munka mennyiségével összhangban legyen (összes óraszám = kredit*30 óra).

4. sor: Legfeljebb 3, már korábban hallgatott tárgy adható meg a 4.1 sorban. A 4.2 és 4.3 sorok *vagyilag* lehetőségek megadására szolgálnak, például abban az esetben, ha az egyik tárgynak korábban oktatott változatai is megfelelnek. **5. sor:** A *kizáró tantárgyaknál* azokat a tárgyakat kell felsorolni, amelyek tematikái a most akkreditálandó tárggyal 75% vagy annál nagyobb átfedést mutatnak.

6-7. sorok: A felelős tanszék és oktató hatáskörét, ill. kijelölésének feltételeit a *Képzési Kódex 2001* c. dokumentum 9.1 fejezete tartalmazza.

Tematika				
7.	A tantárgy az alábbi témakörök ismeretére épít			
	Nyelvek és automaták, programozási alapismeretek, minimális valószínűségszámítás			
8.	A tantárgy célkitűzése, feladata a szakképzés céljának megvalósításában			
	Választható tárgy MSc hallgatóknak			
9.	A tantárgy részletes tematikája			
	A természetes nyelvek, főként a magyar feldolgozásának problémái első sorban mérnöki, megvalósíthatósági szempontból.			
	1. A számítógépes nyelvészet helye, feladata.			
	2. Nyelvtörténet, nyelvosztályozás			
	3. Karakterek, kódolások. Betűstatisztikák, és hasznuk, Shannon-féle nyelvazonosítás.			
	4. Sorrendezési algoritmusok – gyors betűfák: béta- és gammatree.			
	5. A nyelvi statisztikák megbízhatósági kérdései és korlátai.			
	6. Szótan – szavak, szóalakok. Miből mennyi van? Aktív és passzív nyelv			
	7. Generatív modellek morfológiára. Generatív modelltől az gyors analízisig – hatékony szóanalízis eszközei (Ispell, Humor, véges és több szintű automaták, Frey algoritmus). Hely- és időigény viszonya az implementációkban			
	8. Statisztikai módszerek helyesírás-ellenőrzésnél, korrekciónál és generálásnál. (n-gramok, Markov láncok, shake and bake)			
	9. Szószintű helyesírás-ellenőrzés minősítése, tesztelése			
	10. Mérhető-e a mérhetetlen – a helyesírás-ellenőrzés minőségének elvi és gyakorlati korlátja			
	11. Többértelműség a szavak szintjén, annak kiküszöbölése.			
	12. Szóalaktan alkalmazásai – szótárak, intelligens keresés...			
10.	Követelmények, az osztályzat (aláírás) kialakításának módja			
	szorgalmi időszakban	Részvétel az órákon. Mintaprogram készítése. Beszámoló egy választott témából.	vizsgaidőszakban	Szóbeli vizsga.
11.	Pótlási lehetőségek			
	A TVSz-ben leírtak szerint			
12.	Konzultációs lehetőségek			
	Igény szerint			
13.	Jegyzet, tankönyv, felhasználható irodalom			
	Farkas Ernő/Naszódi Mátyás: Magyar nyelv mondatok elemzése természetes nyelv interfész céljából. SzTAKI-kiadvány, 1993			
	Naszódi Mátyás: Saját kiadatlan 40 oldalas jegyzet			
	Több magyar és angol nyelvű témabavágó cikk			

14.	A tantárgy elvégzéséhez átlagosan szükséges tanulmányi munka mennyisége órákban (a teljes szemeszterre számítva)		
	14.1	Kontakt óra	28
	14.2	Félévközi felkészülés órákra	12
	14.3	Felkészülés zárthelyire	0
	14.4	Zárthelyik megírása	0
	14.5	Házi feladat elkészítése	0
	14.6	Kijelölt írásos tananyag elsajátítása (beszámoló)	20
	14.7	Mintaprogram készítés, tesztelés	15
	14.8	Vizsgafelkészülés	15
	14.9	Összesen	90
15.	Ellenrz adat		Kredit * 30

A tantárgy tematikáját kidolgozta			
16.	Név	beosztás	Munkahely (tanszék, kutatóintézet stb.)
	Naszódi Mátyás	tanársegéd	Volt Sztaki, Szit(Vik), Morphologic

A tanszékvezet		
17.	Neve	aláírása
	Dr. Rónyai Lajos	

Megjegyzések

14.1 sor: Értéke automatikusan kitöltődik az rlap elektronikus változatában, a „Követelmény” címszónál megadott óraszám értékek alapján, az (eladás+gyakorlat+labor) * (14 oktatási hét) formula szerint. **14.4 sor:** Értéke 0, ha a zárthelyik íratása kontakt órákon történik, egyébként pedig a minimálisan szükséges számú zárthelyi megírásához felhasználható idő (a pót zárthelyik nélkül). **14.7 sor:** Az „Egyéb elfoglaltság” szöveg helyére a tevékenység konkrét megnevezését kell írni.

15. sor: Az itt szereplő értéknek és a **14.9 sorban** automatikusan megjelenő tanulmányi óraszám összegnek hozzávetőlegesen meg kell egyeznie! Tájékoztatásul azt vegyük figyelembe, hogy a hallgatók által egy szemeszterben átlagosan 30 kreditnyi munkamennyiséget kell teljesíteni, azaz a szorgalmi és vizsgaidőszak során elvárt terhelés összesen kb. 900 munkaóra.