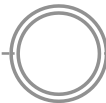




A nyelvtechnológia alapjai



Dömötör Andrea

domotor.andrea@itk.ppke.hu

Yang Zijian Győző

yang.zijian.gyozo@itk.ppke.hu



Dömötör Andrea, Yang Zijian Győző

Szoba: 314

E-mail:

domotor.andrea@itk.ppke.hu

yang.zijian.gyozo@itk.ppke.hu

Honlap: <http://users.itk.ppke.hu/~yanzigy/nyelvtech/>



Csoport	Terem	Gyakorlat
Hétfő ENY	220 PC labor	Hétfő: 08:15 - 10:00
Hétfő ITK	219 PC labor	
Kedd	219 PC labor	Kedd: 08:15 - 10:00
Szerda	322 PC labor	Szerda: 08:15 - 10:00

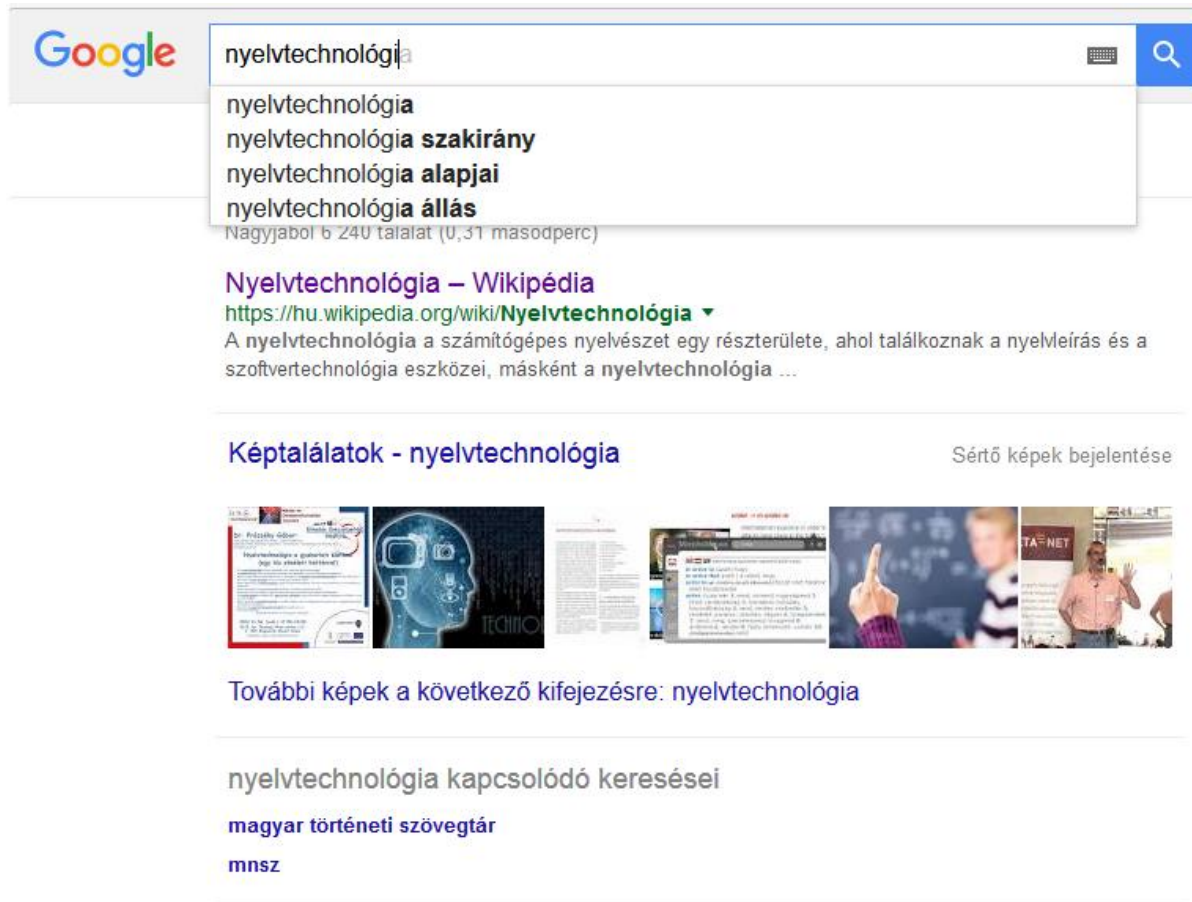
- Maximum 3 hiányzás.
- 2 kötelező nagy házi feladat
- Minden ZH és nagy HF: min 50%

1. ZH	2. ZH	Pót ZH
2019. október 21.	2019. december 2.	2019. december 10.
2019. október 22.	2019. december 3.	
2019. október 23.	2019. december 4.	



NYELVTECHNOLÓGIA

Hol találkozunk vele?





Gépi fordítás

[Bejelentkezés](#)

FordítóAzonnali fordítás kikapcsolása

angol magyar **kínai** Nyelvfelismerés ▾

↔ magyar angol kínai (egyszerűsített) ▾ **Fordítás**

语言技术

Ä 🔊 拼 ▾

Yǔyán jìshù

Nyelvtechnológia

Nyelvtechnológia
nyelvi technológiában
a nyelv technológia
Fordítás javítása
Nyelvtechnológia **Alkalmaz**

 **Módosítási javaslat**

A helyesírás támogatása - Word

FÁJL KEZDŐLAP BESZÚRÁS TERVEZÉS LAPELRENDEZÉS HIVATKOZÁS LEVELEZÉS VÉLEMÉNYEZÉS NÉZET Győző Ya...

Beillesztés Vágólap Betűtípus Bekezdés Stílusok Szerkesztés

Calibri (Szövegtörzs) 11

A helyesírás támogatása

A nyelvtudományi technológiának talán az egyik leghasznosabb törekvése a gépi fordítás, a fordítástámogató programok fejlesztése mellett a helyesírási tanácsadás. Ehhez persze nem ülhet ott a számítógép előtt egy nyelvész, aki állandóan online van, és válaszol mindenkinek minden helyesírási kérdésére, mint valami call-centeres dolgozó, így ebben a témakörben is a gépesítés volt és lesz is a megoldás kulcsa, ám ez jelentős nyelvtudományi fejlesztést igényel. A számítógép szemantikai értelmezésére nem számíthatunk a helyesírásban való segítségnyújtás kapcsán, viszont egy végeláthatatlan hosszúságú korpuszt is képtelenség lenne létrehozni ahhoz hogy minden szónál a gép végigpörgesse az egész listát, hogy megtalálja-e benne, és hogy heylesen írta-e le az illető az adott kifejezést. Az is egy életen át tartó munka lenne, hogy valakik elkészítsék ezt a korpuszt, és még abban sem lennénk biztosak, hogy nem maradt-e ki belőle valami. Így a magyar helyesíráshoz figyelembe kell vennünk a három helyesírási alapelvet, vagyis a hangzás szerinti ejtés, elvét. A sztenderd köznyelvtől való eltéréseket, mint a nyelvvelhetjük figyelembe, főként azért sem, mert ezek listája folyamatosan bővül, és egységesek, sőt olyan is előfordulhat, amit nem is biztos, hogy beszélve az idegen kifejezések magyarításairól.

magyar magzat Az összes kihagyása Felvétel a szótárba Hiperhivatkozás... Új megjegyzés

OLDAL: 1 / 1 203 SZÓ MAGYAR 100%

„Intelligens” kérdés-válasz rendszerek











„Célja megvalósítani a kommunikációt ember és számítógép között **természetes** nyelven, továbbá megalkotni az ehhez szükséges számítógépes programokat.”



- A keresőszóval azonos tartalmú kifejezések keresése

Keresőszó:

alma

Találja meg:

almaként
almák
vadalma

Ne találja meg:

almanach
almárium
hatalma

- Több szavas kifejezések

házi feladat, száj- és körömfájás, Petőfi Sándor

- Szinonimák

kutya



ebadó

- Szavak szintjén
 - A szavak végén az i legtöbbször rövid.
 - A főnevek végén az ú, ű általában hosszú, de vannak kivételek.

- Szavak szintjén
 - A szavak végén az i **legtöbbször** rövid.
 - A főnevek végén az ú, ű **általában** hosszú, de vannak **kivételek**.



- Szavak szintjén
 - A szavak végén az i **legtöbbször** rövid.
 - A főnevek végén az ú, ű **általában** hosszú, de vannak **kivételek**.
- Mondatok szintjén
 - Alany: az a mondatrész, amellyel valakiről vagy valamiről állítunk valamit.

- Szavak szintjén
 - A szavak végén az i **legtöbbször** rövid.
 - A főnevek végén az ú, ű **általában** hosszú, de vannak **kivételek**.
- Mondatok szintjén
 - Alany: az a mondatrész, amellyel **valakiről** vagy **valamiről** állítunk **valamit**.



- Szavak szintjén
 - A szavak végén az i legtöbbször rövid.
 - A főnevek végén az ú, ű általában hosszú, de vannak kivételek.
- Mondatok szintjén
 - Alany: az a mondatrész, amellyel valakiről vagy valamiről állítunk valamit.
- Jelentés, világismeret
 - *Egyik nap még a nap is sütött.*
 - *Péter megvette Az ember tragédiáját.*
 - *Nem én leszek az első amerikai elnök, aki háborút veszít.*
 - *Lujza éhes volt. Bement egy étterembe. Spagettit és kólát rendelt. A pincér gyorsan ki is hozta neki. Lujza elégedett volt, ezért amikor végezett, jó sok borravalót hagyott ott.*



- Szövegszegmentáció
 - Mondatokra bontás (hol van a mondathatár?)
 - Tokenizálás (我喜欢语言技术。 – 我 喜欢 语言技术 。)



- Szövegszegmentáció
- Lemmatizálás (szótő - toldalék)
 - kölyköt – kölyök, jött – jön

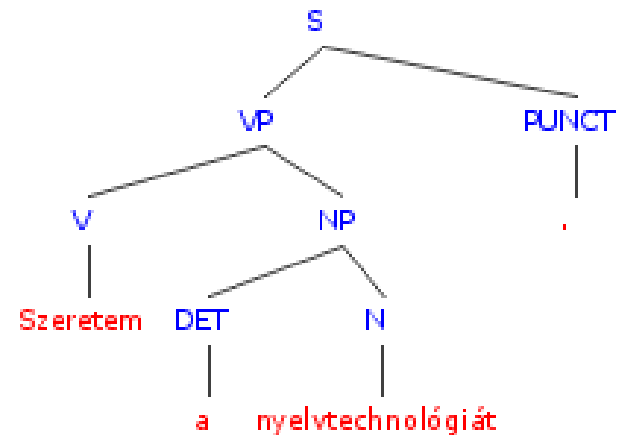


- Szövegszegmentáció
- Lemmatizálás (szótő - toldalék)
- Szófaji elemzés
 - kölyköt#kölyök#[FN], jött#jön#[IGE]



- Szövegszegmentáció
- Lemmatizálás (szótő - toldalék)
- Szófaji elemzés
- Morfológia
 - kölyköt#kölyök#[FN][ACC], jött#jön#[IGE][Me3]

- Szövegszegmentáció
- Lemmatizálás (szótő - toldalék)
- Szófaji elemzés
- Morfológia
- Szintaktikai elemzés
 - Szeretem a nyelvtechnológiát.





- Lexikai szemantika
 - Leült egy ...



- Lexikai szemantika
- Névelemfelismerés
 - Pázmány Péter Katolikus Egyetem



- Lexikai szemantika
- Névelemfelismerés
- Többértelműség
 - vár, körte, nap



- Lexikai szemantika
- Névelemfelismerés
- Többértelműség
- Tartalom kivonatolása



- Lexikai szemantika
- Névelemfelismerés
- Többértelműség
- Tartalom kivonatolása
- Szentiment analízis



- Lexikai szemantika
- Névelemfelismerés
- Többértelműség
- Tartalom kivonatolása
- Szentiment analízis
- Terminológia

- Döntéstámogató rendszerek



- Döntéstámogató rendszerek
- Kérdés-válasz rendszer
 - chatbotok
 - intelligens személyi asszisztens

- Döntéstámogató rendszerek
- Kérdés-válasz rendszer
- Gépi fordítás

Hol lehet hasznosítani ezt a tudást?

Kutatás



Magyar cégek



Globalese



Nemzetközi cégek





1. Karakterkódolás; Lokalizáció
2. Dokumentumformátumok; XML
3. XML: Schema, CSS, XSLT
4. XML: SAX, DOM
5. Kétszintes morfológia
6. Gyakorlati morfológia
7. Szintaktikai elemzés
8. Szemantika; Szemantikus hálók; Ontológia
9. Gépi fordítás
10. Kiértékelési módszerek
11. Neurális háló a nyelvtechnológiában; Word embedding



Köszönöm a figyelmet



ध्यान देने के लिए आपका धन्यवाद
आप अपना ध्यान के लिए धन्यवाद
थैंक यू फॉर योर अटेंशन

Thank you for your attention
Thank you for your attention
Thank you for the attention

Merci de votre attention
Je vous remercie pour votre attention
Merci pour l'attention

感谢您的关注
谢谢你的关注
谢谢您的注意