



Nyelvészet az informatikában – informatika a nyelvészetben: Bevezetés, követelmények ismertetése

Vincze Veronika

vinczev@inf.u-szeged.hu

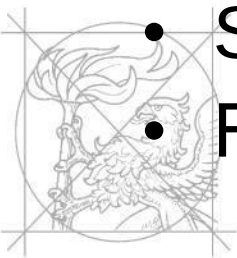
Bevezetés

- **Informatika / számítástudomány**
- **Nyelvészet / nyelvtudomány**
- **Mivel foglalkoznak?**
- **Milyen tágabb tudományterületbe sorolandók?**



Informatika

- A számítógép-tudomány vagy számítástudomány a matematika egyik, igen fiatal tudományága, amely a számítógépek tervezésének és működtetésének elméleti, matematikai alapjaival foglalkozik (Wikipédia)
- Computer [science](#)
- Matematika / természettudomány
- SZTE TTIK ☺
- Földrajz, fizika, matematika, biológia...



Nyelvészet

- A nyelvészet az emberi nyelvekkel foglalkozó tudományág (Wikipédia)
- Linguistics
- Bölcsészettudomány
- SZTE BTK
- Történelem, irodalom, filozófia...



-> hol ér össze a kettő?

Nyelvészet + informatika

- **Nyelv- és beszédtechnológia:**
 - írott és a hangzó nyelv feldolgozása
 - nyelvi produktumok előállítása
- **Natural language processing (természetesnyelv-feldolgozás)**
- **Computational linguistics (számítógépes nyelvészet)**
- **Human language technology (nyelvtechnológia)**



Célok

- az ember-ember, az ember-gép kommunikáció hatékonyabbá tétele
- az emberi munkavégzés megkönnyítése újszerű, számítógépes termékek és szolgáltatások biztosításával
- hátrányos helyzetű csoportok (siketek, gyengénlátók, baleset következtében beszédképességüket elvesztők, idegen nyelveket nem tudók) életminőségének javítása





Határterületek

- Nyelvészet
- Lexikográfia
- Szoftvertechnológia
- Pszichológia
- Matematika
- Informatika
- Fizika
- Fiziológia
- Neurológia
- Biológia
- ...

Nyelvtechnológia a mindennapokban

- Helyesírás-ellenőrző
- Webes keresés (Google)
- Fordítóoldalak (Google Translate, webfordítás)
- Címkézés [origo]
- Hanghívás
- Telefonos tudakozó

...

vö.: <http://www.inf.u-szeged.hu/rgai/nlp>



A nyelvészet mint természettudomány

- **Módszertan**
- **Adatok gyűjtése: megfigyelt nyelvi jelenségek**
- **Hipotézis felállítása**
- **Hipotézis igazolása / cáfolata adatokkal**
- **Modellalkotás**
- **Lásd: biológiai, kémiai, fizikai kísérletek + matematikai tételek bizonyítása**



Példa

- Hipotézis: magyarban a birtok többes számát a birtokon, a birtokosét a birtokoson jelöljük
- Adatok (korpuszból v. anyanyelvi beszélőktől):
 - *a fiú háza – a fiúk háza – a fiú házai – a fiúk házai*
 - *az ő háza – az ő házuk (*az ők háza) – az ő házai – az ő házaik (*az ők házai)*
- Hipotézis elvetése / finomítása
- A modellben külön kell kezelni a névmási birtokost



A kurzus célja

- Hogyan segítheti a (természettudományi szemléletű) nyelvtudomány az informatikát (természetesnyelv-feldolgozást)?
- A különböző nyelvészeti részterületek hogyan képeződnek le a természetesnyelv-feldolgozás gyakorlatában?
- Informatikai eszközökkel és technológiákkal hogyan lehetséges a nyelvészek munkáját segíteni?



Tematika

- Nyelvészeti alapfogalmak
- Fonetika, fonológia
- Morfológia
- Szintaxis
- Szemantika
- Lexikográfia
- Tipológia
- Korpuszok és adatbázisok
- Nyelvészeti adatgyűjtés



Követelmények

- 2, az órák anyagához kapcsolódó rövid házi feladat elkészítése (szorgalmi időszakban)
- kollokvium – írásbeli vizsga (vizsgaidőszakban)





Menetrend

- **Szeptember 11-én, 18-án és október 16-án elmarad az óra!**
- **Pótlás később megbeszélendő időpontban**
- **Házi feladatok beadása: október vége, illetve november közepe (figyeld a honlapot!)**
- **December 4.: elővizsga**

Elérhetőség

A kurzus anyaga elérhető:

<http://www.inf.u-szeged.hu/~vinczev/oktatas.html>

**MTA-SZTE Mesterséges Intelligencia
Kutatócsoport**

6720 Szeged, Tisza Lajos krt. 103.
45/b szoba

Telefon: +36 62 544 140

Fogadóóra: előzetes egyeztetés alapján

