

# Önálló labor feladatkiírásaim – 2016. tavasz

*(ezekhez kapcsolódó saját témával is megkereshetnek)*

*Mészáros Tamás*  
<http://www.mit.bme.hu/~meszaros/>

Budapesti Műszaki és Gazdaságtudományi Egyetem  
Méréstechnika és Információs Rendszerek Tanszék

# Információkeresés és -szolgáltatás, intelligens ágensek

- Információkeresés (information retrieval)
  - tárolt szöveghalmazban az igényelt információ megtalálása
  - szövegek elemzése és keresések futtatása
  - strukturált szövegek (XML) létrehozása, feldolgozása és felhasználása  
pl.: webes keresőrendszerek, könyvtári keresők, üzleti adatelemző és -tisztító rendszerek, spam szűrés, desktop kereső, tudáskinyerők, stb.
- Információszoolgáltatás (kapcsolat a felhasználóval)
  - felhasználó modellezés (az igényelt információ kontextusa)
  - információbevitel, keresés és lekérdezés természetes nyelven
- Intelligens ágensek alkalmazása
  - észleli a környezetét és beavatkozói segítségével önállóan cselekszik
  - kommunikál más entitásokkal (ágensekkel, emberekkel, stb.)
  - közösségi entitás (több-ágens rendszerekben kooperál és verseng)
  - modellezési és alkalmazásfejlesztési eszköz (Java-alapú)  
pl.: web indexelő robot, szövegelemző hálózat, felhasználói interfész ágens, ...

# Kontrollált természetes nyelvek alkalmazása

- Természetes nyelvű kommunikáció a számítógéppel
- Kontrollált természetes nyelvek
  - Mintha természetes lenne (jól érthető, megtanulható)
  - Alkalmazási területre szabható (mesterséges)
  - Jól elemezhető (igen egyszerű elemző is elég lehet)
  - Jól értelmezhető (egyértelmű szemantika, automatikus fordítás)
- Mire használható?
  - Bonyolult interfészek egyszerűvé válnak:
    - a nyelv biztosítja az összetettségüket, nem a felületi elemek
    - Olyanok is használhatják, akik egy számítógépes felületen elvesznek
    - Akkor is működik, ha nem áll rendelkezésre megfelelő interfész eszköz
    - Hangalapú kommunikációt is lehetővé tesz (pl. Google Speech Input)
  - Tudásbevitel
    - Elemezhető, a számítógép által megérthető szöveg
    - Formális (pl. logikai) reprezentációra alakítható

# Tudásalapú szövegfeldolgozás és -reprezentáció

- Szövegekkel kapcsolatos tudás felderítése
  - távtartó olvasás (distant reading): pl. statisztikai jellemzők feltárása
  - alapos olvasás (close reading): pl. entitásfelismerés, strukturált elemzés
  - metaadatok (HTML microformats, RDFa, stb.)
  - emberi szakértők elemzései, annotációi (kontrollált nyelvű bevitellel)
- reprezentációja
  - strukturált szövegformátumok (TEI XML)
  - tudásbázisok, ontológiák (RDF, OWL)
  - adatkapcsolati hálók (Linked Open Data)
  - adatbázisok (numerikus jellemzők)
- és alkalmazása
  - (jelenleg) irodalmi és történeti művek tudásalapú feldolgozása
  - (elképzelhető) web, közösségi hálók, szakirodalmak
- egy webes kutatói rendszerben
  - felhőalapú tárolás (owncloud, mongodb) és adatmegosztás
  - statisztikai programok (R Studio) és automatizált elemzők (ágensek)

# Irodalmi szövegek számítógépes elemzése

(Az MTA Irodalomtudományi Intézettel közös feladatkiírás)

- Vajon Shakespeare írta az összes művét?
- Ki kivel állt kapcsolatban, kinek mely más szerzőre volt hatása
- Hogyan változott egyes szavak gyakorisága az idő folyamán?
  - 1800-1970 között a „nő” szó a „férfi”-hez képest elenyésző gyakoriságú volt, 1980 óta nagyjából egyforma arányban fordulnak elő az angol irodalomban.
  - Az „1880” szó használatának gyakorisága 1912-re feleződött meg, míg az „1973” már nagyjából 1983-ra elérte ezt a szintet. Egyre gyorsabban felejtünk?
  - Az 1800-1840 között találmányok nevei kb. 66 év után terjedtek el írásban, míg az 1880-1920 közöttieknek ez csak 27 évig tartott
- 2010 óta jelentek meg jelentősebb (angol) publikációk a témakörben
- A magyar írásbeliség ilyen jellegű vizsgálata úttörő munkának számít
  - Példaként Mikes Kelemen igen terjedelmes életművének vizsgálata a cél
  - Digitalizált változatban már rendelkezésre áll, most zajlik a szótárkészítés
  - A szövegnek létezik mai átírása, de az eredeti nyelvezet is vizsgálható
  - A szótár a szavak értelmezésében, eltérő szóalakok felismerésében segít
  - A feladat nyitott, egyéni ötletekkel is elő szabad, sőt kell állni

# Általános tudnivalók – mit várok és mit nyújtok

- „Mindent szabad, ami örömet okoz”
  - a motiváció érdekében a feladatkírást a hallgatóval közösen véglegesítem
  - szabad saját (akár irreálisnak tűnő) ötletekkel változtatni a feladaton
  - a feladat méretének helyes meghatározása a konzulens feladata
- Előzetes jelentkezés a konzulensnél
  - rövid bemutatkozás (előismeretek)
  - miért érdekli a téma (mi a motiváció)
  - milyen elképzelései vannak a feladatkíírás módosításával kapcsolatban
- Az önálló labor menete
  - közösen megalkotott specifikáció és vázlatos munkaterv
  - heti rendszeres konzultációk (aki jól halad, annál ritkábban is lehet)
  - fontos a terület önálló felfedezése, de irodalmat bőségesen adok
  - a terveket megbeszéljük, az implementáció önálló munka
  - nem a tökéletes termék a cél, hanem a terület megismerése, megértése