

# Kontrollált természetes nyelvű lekérdezés webes adatbázisokhoz

Mészáros Tamás, BME MIT

Dobrowiecki Tadeusz, BME MIT

Kiss Margit, ELTE BTK

# Bevezetés helyett

Főnévi vonzattár

Felhasználó: Mészáros Tamás [Portál admin] (Kijelentkezés)  
Nyomtatható változat

[Információk](#)

[Keresés főnevek és vonzatok szerint](#)

[Publikációk](#)

Navigáció: [Főnévi vonzattár](#) - [kérdés a főnévlistához](#)

## Keresés a főnévi vonzattárban kérdés segítségével

Az alábbi mezőben magyar nyelven lehet feltenni főnevekkel és vonzataikkal kapcsolatos kérdéseket.  
A rendszer csak bizonyos szerkezetű kérdéseket tud megválaszolni, ezért kontrollálja a kérdés feltevését.

**Első használat előtt érdemes elolvasni a [prediktív szövegbevitellel kapcsolatos tájékoztatót](#).**

Kérdés:

DEBUG:



# A feladat

- n Űrlapok helyett természetes nyelvű lekérdezés („emberközelibb” felhasználói felület)
- n „Megérteni”, amit a felhasználó ír (szintaktikai és szemantikai elemzés)
- n Lefordítani gépi lekérdezéssé (például SQL vagy XPath kifejezéssé)



# Miért érdemes?

- n A felhasználó saját természetes közegében fogalmazza meg az információs igényét.
- n Jobban érthető és használható felhasználói felület (űrlapok helyett)
- n Az egyszerűsége ellenére nagyon erős kifejezőerővel bír
- n Rugalmasabb információ-elérést (és felhasználói felületet) tehet lehetővé



# Miért nehéz?

- n A természetes nyelvű szövegek gépi elemzése nehéz feladat
  - .. a szintaktikai elemzés sem könnyű (időben és térben változó nyelv)
  - .. a jelentés meghatározása még nehezebb (a jelentés átadása, egyértelműségi problémák, jelentős háttértudás beépítése szükséges)

 Jelenleg nem oldható meg az általános feladat



# Megoldás: kontrollált nyelv

- n A természetes nyelv használatát korlátozza:
  - .. rögzített nyelvtani szabályok
  - .. rögzített szókincs
- n Két problémát kezelünk egyszerre:
  - .. a szöveg biztosan elemezhető lesz
  - .. egyértelmű jelentés rendelhető hozzá
- n A módszer alkalmazása során:
  - .. a nyelvet létre kell hozni
  - .. a felhasználónak a korlátozott nyelvet kell használnia



# A kontrollált nyelv alkalmazása

- n A feladathoz kell szabni a nyelvet (elégséges nyelvtannal és szókinccsel)
- n A felhasználótól megköveteljük a korlátozott nyelvtan és szókinccs használatát
  - A felhasználót fel kell készíteni a nyelv helyes alkalmazására (oktatás és gyakorlás), vagy...
  - kontrollálni kell a szövegbevitelt
- n Választott megoldás:
  - prediktív szövegbevitel



# A szövegbevitel megvalósítása

- n A felhasználói felületen csak a kontrollált nyelv használható (kötött szerkezetek és szókincs)
- n Folyamatosan korrigálni kell a hibákat (nyelvtani és helyesírási)
- n Segíteni kell a helyes mondatok megfogalmazását



„Folyamatos” nyelvi elemzés és generálás





# Elemzés és generálás

- n A feladat eltér a hagyományos nyelvi elemzéstől
  - .. mondattöredékek hatékony elemzése
  - .. lehetséges bővítések generálása
- n A két feladatot egy rendszerbe ötvözve kell hatékonyan megoldani
  - .. rendezett tömörített erdő reprezentáció
  - .. hatékony és támogatja a generálást

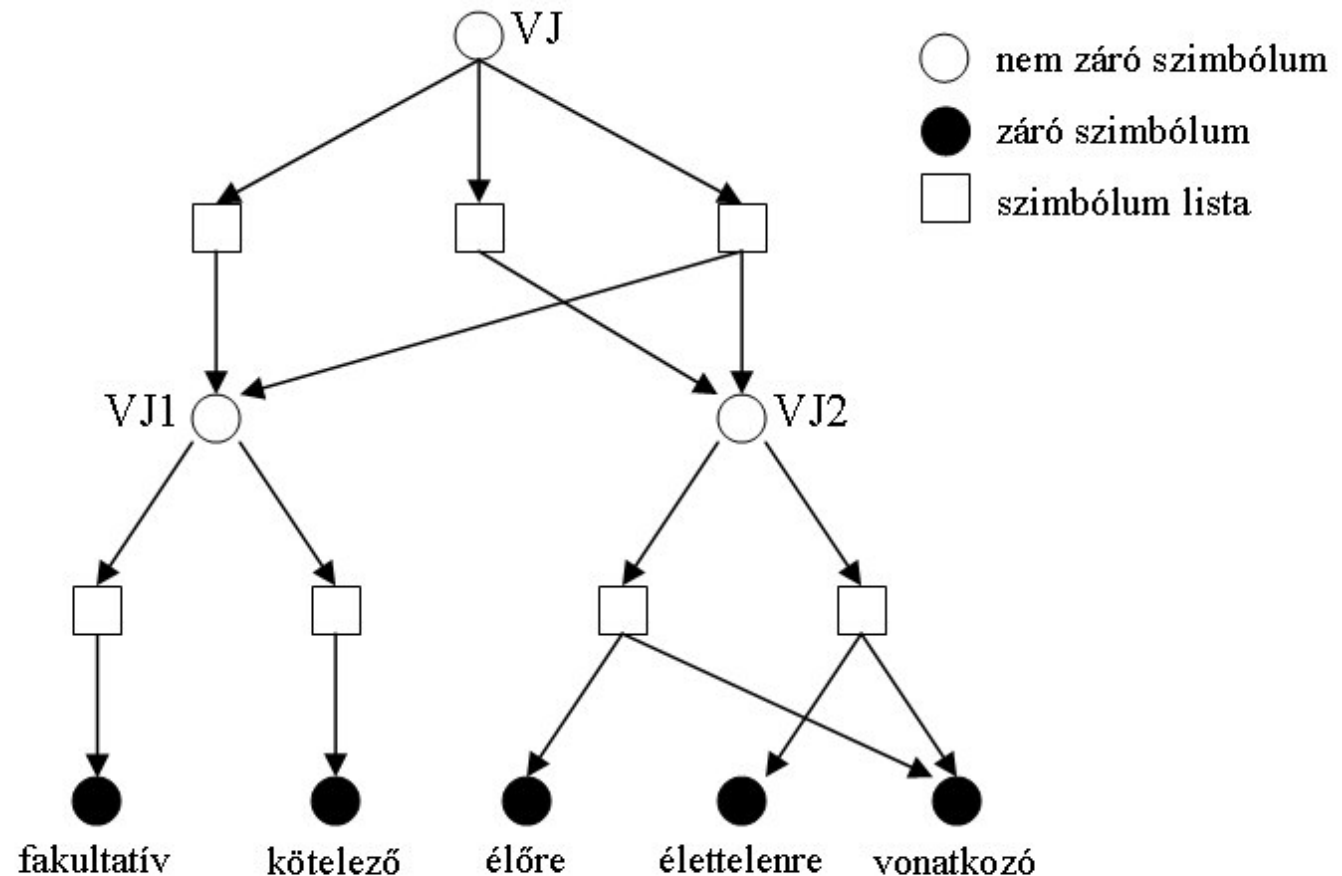


# Rendezett tömörített erdő

- n Nyelvtan és szókincs reprezentáció egy speciális irányított gráffal
- n A nyelvtani (helyettesítési) szabályok különálló fáit egy összevont gráfban reprezentálja (erdő)
- n Polinomiális tár és idő komplexitású (az exponenciális fák helyett)
- n A rendezés segítségével a generálási feladatokra is alkalmassá válik (további részletek a cikkben)

# A rendezett tömörített erdő (példa)

$VJ = VJ1 \mid VJ2 \mid VJ1 VJ2$   
 $VJ1 = \text{fakultatív} \mid \text{kötelező}$   
 $VJ2 = \text{élőre vonatkozó} \mid$   
 $\text{élettelenre vonatkozó}$



3. ábra: a vonzatjelzők (VJ) reprezentációja tömörített erdővel



# Megvalósítás webrendszerekben

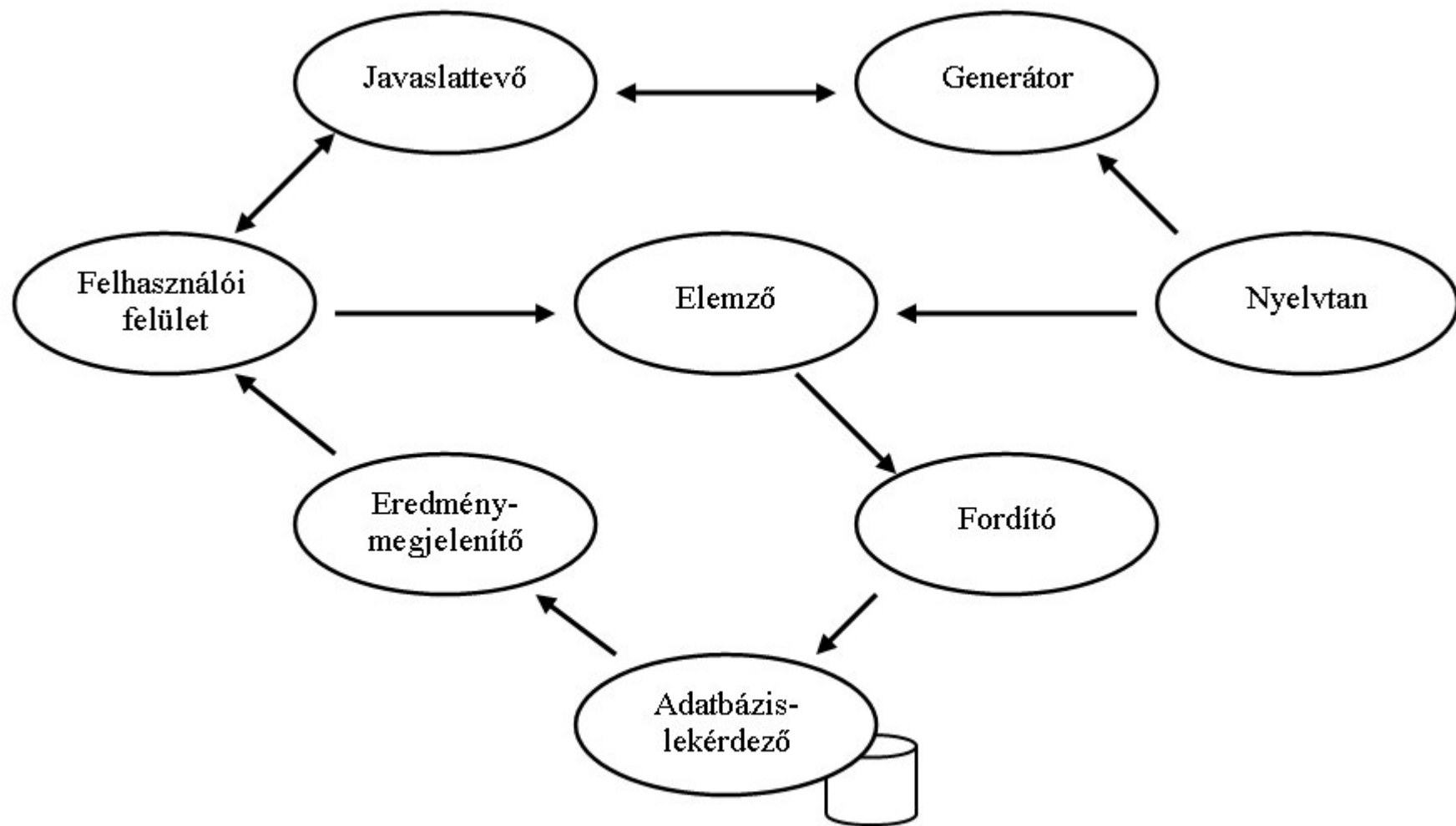
## n Feladatok:

- .. Kontrollált szövegbeviteli felület
- .. Folyamatos elemzés és generálás
- .. A lekérdezések elemzése és fordítása

## n A megoldás építőelemei:

- .. A prediktív szövegbevitelhez: AJAX
  - n Javascript-alapú ellenőrzés a kliensen
  - n Aszinkron XML üzenetek a kliens és a szerver között
- .. Szerver oldali elemzés és fordítás

# A rendszer elvi felépítése



2. ábra: a rendszer főbb komponensei és azok kapcsolódása



# Alkalmazás: főnévi vonzattár

- n Mintegy 31 000 főnév vonzatait tárolja
- n Egy főnévhez harmincféle vonat társulhat
- n A vonzatok között kapcsolatok is lehetnek
- n A hagyományos lekérdezés egy többlépcsős bonyolult űrlaprendszer
- n Különösen a feltételek összekapcsolása jelentett gondot a felhasználók számára
- n Az űrlapokat váltotta fel egy magyar nyelvű kontrollált felhasználói felület

# Keresés a főnévi vonzattárban

Főnévi vonzattár

Felhasználó: Mészáros Tamás [Portál admin] (Kijelentkezés)  
Nyomtatható változat

[Információk](#)

[Keresés főnevek és vonzatok szerint](#)

[Publikációk](#)

Navigáció: [Főnévi vonzattár](#) - [kérdés a főnévlistához](#)

## Keresés a főnévi vonzattárban kérdés segítségével

Az alábbi mezőben magyar nyelven lehet feltenni főnevekkel és vonzataikkal kapcsolatos kérdéseket. A rendszer csak bizonyos szerkezetű kérdéseket tud megválaszolni, ezért kontrollálja a kérdés feltevését. **Első használat előtt érdemes elolvasni a [prediktív szövegbevitellel kapcsolatos tájékoztatót](#).**

Kérdés:   DEBUG:

<http://www.webit.hu/fonevlista/?l=askme>



# Összefoglalás

- n A felhasználó (korlátozott) természetes nyelven fogalmazhatja meg az információs igényét
- n A prediktív szövegbevitellel elkerülhető a felhasználók betanítása
- n Kidolgoztunk egy hatékony elemzési és generálási reprezentációt és algoritmust
- n Kifejlesztettünk egy prototípus webrendszert az AJAX technológia felhasználásával





# Továbbfejlesztési lehetőségek

- n Bonyolultabb nyelvtanok vizsgálata
- n Automatikus nyelvtan generálás ontológiából
- n A fordítás továbbfejlesztése
- n Nyelv és fordító általános XML lekérdezésekhez
- n Integráció a szemantikus webbel (RDF, OWL)
- n Nem csak a weben alkalmazható



# A szerzők elérhetőségei

## **Mészáros Tamás**

BME Méréstechnika és Információs Rendszerek Tanszék

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

[meszaros@mit.bme.hu](mailto:meszaros@mit.bme.hu)

## **Dobrowiecki Tadeusz**

BME Méréstechnika és Információs Rendszerek Tanszék

<http://www.mit.bme.hu/~tade/>

[dobrowiecki@mit.bme.hu](mailto:dobrowiecki@mit.bme.hu)

## **Kiss Margit**

ELTE BTK Nyelvtudományi Doktori Iskola

[kiss.margit@webit.hu](mailto:kiss.margit@webit.hu)