

Bárházi Eszter

SZTE BTK Nyelvtudományi Doktori Iskola, Elméleti nyelvészet program
MTA SZTAKI Géppel Támogatott Megértés Kutatócsoport
barthazi@sztaki.hu

A kutatás célja:

Panaszleveleket humán feldolgozásra előkészítő rendszer kifejlesztése

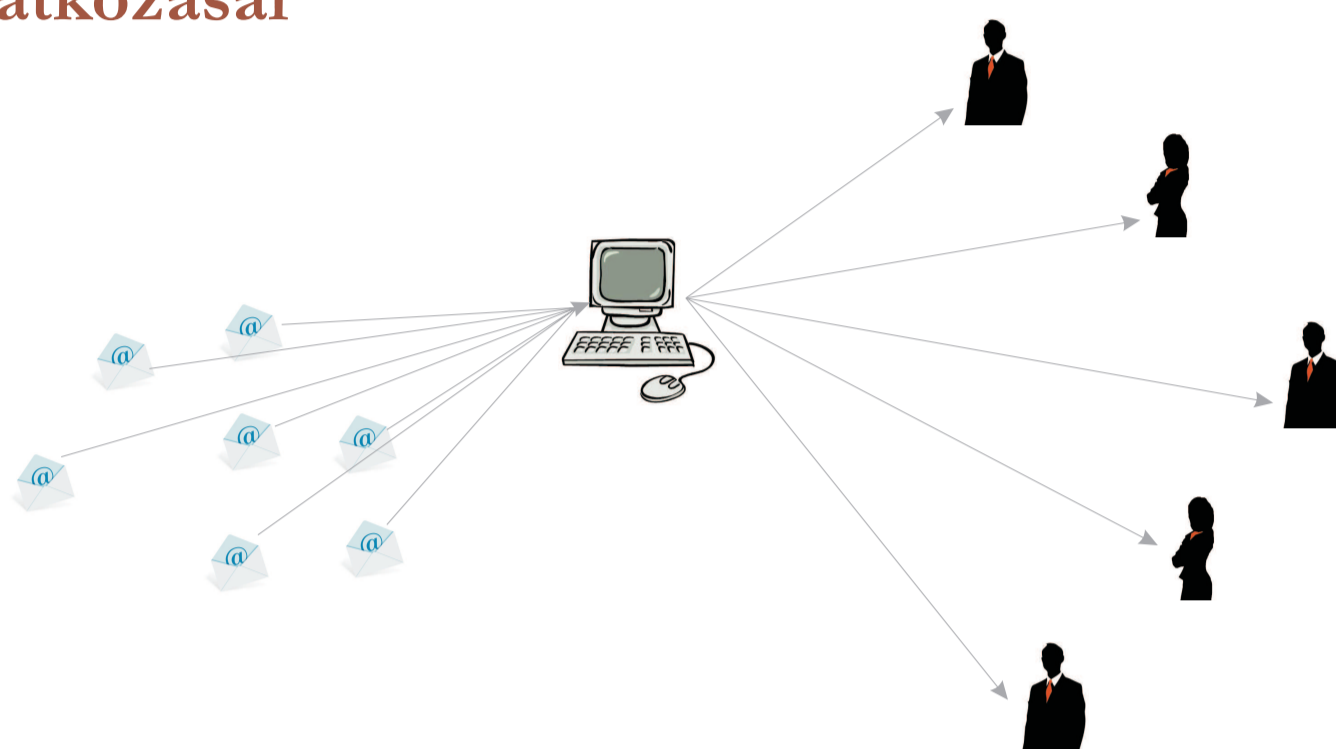
Az előadás váza:

1. A feladat és nyelvészeti pragmatikai vonatkozása
2. A szerkezeti egységek azonosítása
3. A kategorizálási feladat
4. Információkinyerés

1. A feladat és nyelvészeti pragmatikai vonatkozásai

Automatikus kategorizáló rendszer:

- megállapítja a beérkező levél panaszkategóriáját
- az előfeldolgozott panaszleveleket egy az adott témában jártas ügyintézőhöz továbbítja



A korpuszra jellemző:

- Hétköznapi nyelvhasználat
- Az egyes szakterületek terminusainak használata rendszertelen, pontatlan
- Irreleváns információk
- Zavaros megfogalmazás

* 29. levél
* Tisztelt Minisztérium A nevem Person édesanyám neve Name Apám Name volt.
* Az édesanyám itt halt meg Magyarországon 28 éves volt.
* Sajnos én genetikailag örököltem egy súlyos betegséget erős fájdalom van nekem izom izületi gyulladás mivel gyermekkorom óta depresszióban szenvedek az súlyosítja a betegséget.
* Az immunrendszerem elhagyott 2 éve kímásokok ez felgyorsította a folyamatot
* 67%-os rában vagyok egy közepiskolás fiannal élek önkormányzati belföldben Gyula
* A másik fiam Name Gyula börtönben van ez is nagyon elszomorítja a lelkemet.
* Nem sajnálnai szeretném magamat csak az lenne a kéréssem hogy szólnak a Francia Nagykövetségen ha tudnak segítsenek nekem
* A bűncselekmény Name értesítők hogy beteg vagyok
* En mostoha gyermek vagyok ennek az országnak nem én akarom ide jönni apám hozott ide minket és nem engedte hogy édesanyámmal visszamenjünk
* En édesanyámat emiatt megszakított a szíve
* A férjem 2000-ben rában meghalt
* En mindig balesetesen étem és sokat dolgoztam Angol Néi szabó szakmában és még betanított lakatosként is dolgoztam a Gyulai Mezőgazdasági 10 évig
* Kérem önöket legyenek szívesek megmondani nekik és sajnos nem beszélek a Francia nyelvet apám nem engedte a rossz gyermekkorom miatt nem is lát sokat rá időm
* Apám vadállat módra nevelt bennünket
* Nekem az volt a legnagyobb bűnöm, édesanyáma nagyon hasonlított

TÚL TÁG KONTEXTUS
(túl sok inkonzisztens információ a gépnek)

MEGOLDÁS

- szerkezeti egységek detektálása:
- rendszerezik a gép számára inkonzisztens információkat
 - szűkebb kontextusként szolgálnak a bennük előforduló nyelvi információk interpretálásához,
 - mutatják, h a különböző lényegi információkat a levél mely részében érdemes keresni
 - lehetővé teszik az irreleváns részek felismerését, és így figyelmen kívül hagyását
 - egyszerre top-down és bottom-up információkinyerést tesznek lehetővé

A szerkezeti egységek:

- Megszólítás
- Bemutakozás
- Cél
- Probléma
- Javaslat
- Elismerés
- Egyéb körülmények
- Elvárás
- Köszönet
- Lezárás

131. levél
Tisztelt Kóka János Gazdasági Miniszter! Alulírott | Person | Local alatti lakos szívesen fordulok önnek, hogy 1993-ban Mályban kötöttem egy 10 éves bérelti szerződést egy étteremre. Az éttermet beruháztam akkor már az uniónak megfelelően (közel 4 millió Ft-ot ruháztam be, amit a fiam átvéhetett volna tőlem). Sajnos 8 hónappal tovább nem tudtam folytatni az étteremben a munkámat. 1994 június 8-án a tulajdonos fia feltréte az éttermet, holott én 150 ezer Ft bérelti díjat fizettem neki, de nem volt neki elég a pénz. Ekkor nekem 11 év rendezvényt kellett visszamondanom, az étterem nagyon szépen beindult a polgármesteri hivatalból hozzám jártak vacsorázni, de nem tudtak semmit se segíteni. Annyit tudtak hogy kötelezték arra, hogy 24 óra belül állítsa vissza az eredeti állapotot. Ő ezt nem tette meg, majd megbüntették amit nem is fizetett ki. En bírósághoz fordultam, a part 9 évre rá meg is nyertem. Meggyógyultam én sajnos nagyon beteg lettem, gyógyíthatatlan betegséget szereztem 100%-os rában vagyok, dolgozni az életben már nem tudok, pedig szívesen tenném. Budapesti lakásomból azért kellett elköltözni, mert állandóan fenyegetett pedig ő tartozik nekem. Már sajnos itt is megjelent Name mert az egyik bíróságon a volt szakácsom véletlen elszöktetett a dínünket. Most már ide is jár, ezzel nem lehet semmit se tenni, nem értem szeretném tudomására hozni hogy semmijük sem volt, kerékpárral jártak, majd két hónapra miután engem kirúgtak kiadták az éttermet egy újabb nérlőnek az én berendezéseimmel. Itt volt egy rossz Trabant, Lada típusú gépkocsija, ezt a bérlet is kirúgtam majd kiadta egy harmadik bérletnek. A harmadik bérletből épített egy másik házat az udvaron amibe laknak majd felszerelte magát autókkal. 2002-ben újra eladta a bérelti jogot ebből vettek Name az élettársának egy lakást és egy hajót ami a Balatonon áll. Mivel Name-nak a beállása szerint hatalmas adósságot vallott be a nevére semmi sem lehetett, ezért került a bit. az élettársa Name nevére. 10 évig ha veszteséges volt az étterem és a fitnász terem akkor, hogyan lehetett ennyi mindent vásárolni, mielőtt eltek ha nem tudja kifizetni nekem azt a pénzt amit a bíróság megítélt (jelenleg kb. 20 millió Ft). Jelenleg egy kínai bérlet az éttermet az én beruházásommal, és ő veszi fel a pénzt. Ha ez nem működik akkor mi? Mit érhet az ilyen embert lecsukni és miért nem tudok a pénzemhez juttatni? Kérem a tisztelt kollégáit, hogy segítsenek, hozzanak olyan ítéleteket, hogy a pénzemet megkaphassam és az ilyen csúkok megkaphassák a nekik járó büntetést. Budapest, 2006. július. 16. Köszönettel MSZP pártoló tag U. mellékelve megküldöm a bírósági végzéseimet.

2. A szerkezeti egységek azonosítása

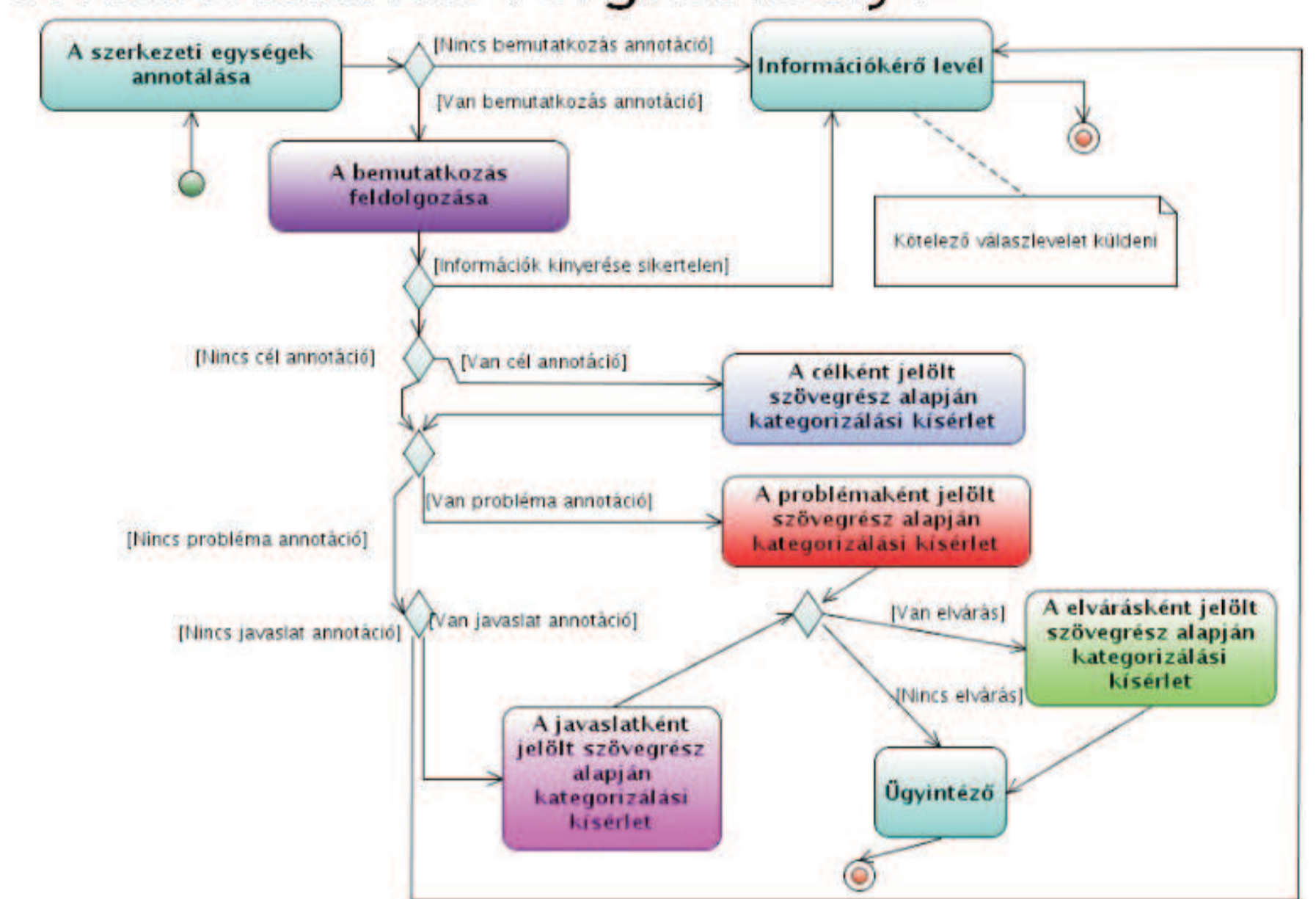
Forrás	Meghatározás	Példa
lexikai jellemzők	kategoriasemleges definitív kifejezések és kategoriaspecifikus kifejezések	Pl.: <i>előre is köszönöm</i> , amely a Köszönet szerkezeti egységre jellemző
szemantikai jellemzők	absztrakt címkék	Pl.: [személynév], [intézménynév] stb
“frázisszerkezeti” jellemzők	a lexikai és a szemantikai jellemzők kombinációi	Pl.: <i>tisztelt</i> [személynév]
levélszerkezeti jellemzők	a SZE-ek egymáshoz képest elhelyezkedésére vonatkozó heurisztikák	Pl.: az Egyéb körülmények szerkezeti egység követni szokta a Probléma szerkezeti egységet
lexikai-levélszerkezeti jellemzők	a lexikai és a levélszerkezeti jellemzőket kombináló heurisztikák	Pl.: levél elején külön sorba álló kifejezés vagy mondat amely kategoriaspecifikus kifejezést tartalmaz, az valószínűsíthetően a Cél SZE-et jelöli

MEGOLDÁS

Kulcsalapú modell (I. Jurafsky 2004) valószínűségi modell: különböző (lexikai, szemantikai és szerkezeti) forrásokból származó információk (kulcsok) együttes figyelembevételével detektálja a szerkezeti egységeket

3. A kategorizálási feladat

A KATEGORIZÁLÁS forgatókönyv



MEGOLDÁS

A KATEGORIZÁLÁS forgatókönyv:
- a panasz kategóriába soroláshoz szükséges lépéseket tartalmazza
- csak a kategorizálás szempontjából releváns szerkezeti egységeken belüli nyelvi információkat veszi figyelembe

4. Információkinyerés

A különböző forrásokból származó kulcsok alapján azonosított szerkezeti egységek segítik a levelek értelmezését, amennyiben kontextusát adják a bennük előforduló nyelvi információknak:

[szervezet] vagy [titulus]+[személynév]+[titulus] vagy [titulus]+[személynév] vagy [személynév]+[titulus] vagy [személynév] vagy [titulus] absztrakt címke alatti kifejezések: meghatározzák a levélben felvázolt esemény szereplőit, ezen belül:

- Bemutakozás SZE-ben: ki a panaszos (valamint az azonosításhoz szükséges anyja neve)
- Probléma SZE-ben: a panaszesemény résztvevői
- Megszólítás SZE-ben: kitől vár a panaszos intézkedést

[dátum]: meghatározza az esemény, illetve részeseemények idejét

- Megszólítás/Lezárás SZE-ben: a levélírási idejét
- Bemutakozás SZE-ben: a panaszos születési adatait)
- Probléma SZE-ben: a panaszesemény, ill. panaszrészeseemények idejét

-kategoriaspecifikus kifejezések, ill. [Törvény] absztrakt címke alatti konkrét törvényre hivatkozás a Cél, Probléma, Javaslat és Elvárás SZE-ben: meghatározzák az adott panaszesemény panaszkategóriáját

Hivatkozás
Jurafsky, D. 2004. Pragmatics and Computational Linguistics. In: Horn, L.R., G. Ward (szerk.) The Handbook of Pragmatics. Cambridge MA: MIT Press. 578-604.

Háttérrodalom
Bunt, H. 2000. Dialogue and context specification. In: Bunt, H., W. Black (szerk.) Abduction, Belief and Context in Dialogue Studies in Computational Pragmatics. Amsterdam: Benjamins. 81-151.
Polanyi, L. 2001. The Linguistic Structure of Discourse. In: Schiffrin, D., D. Tannen, H. E. Hamilton (szerk.) The Handbook of Discourse Analysis. Oxford: Blackwell. 265-281.
Redeker, G. 2000. Text and discourse structure. In: Bunt, H., W. Black (szerk.) Abduction, Belief and Context in Dialogue Studies in Computational Pragmatics. Amsterdam: Benjamins. 233-264.
Schank, R. C. 1986. Explanation Patterns – Understanding Mechanically and Creatively. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
Webber, B. L. 2001. Computational Perspectives on Discourse and Dialog. In: Schiffrin, D., D. Tannen, H. E. Hamilton (szerk.) The Handbook of Discourse Analysis. Oxford: Blackwell. 798-816