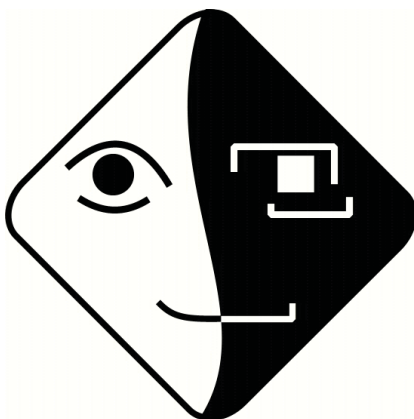


VII. Magyar Számítógépes Nyelvészeti Konferencia



MSZNY 2010

Szerkesztette:

Tanács Attila
Vincze Veronika

Szeged, 2010. december 2-3.
<http://www.inf.u-szeged.hu/mszny2010>

ISBN: 978-963-306-075-9

Szerkesztette: Tanács Attila és Vincze Veronika
{tanacs, vinczev}@inf.u-szeged.hu

Felelős kiadó: Szegedi Tudományegyetem, Informatikai Tanszékcsoport
6720 Szeged, Árpád tér 2.

Nyomtatta: Planet Corp. Szolgáltató Kft.
6771 Szeged, Makai út 4.

Szeged, 2010. november

A konferenciakötet megjelenését az NKTH a TECH_08_A2/2-2008-0092
(MASZEKER) azonosítójú projekt keretében támogatta.



Előszó

2010. december 2-3-én hetedik alkalommal rendezzük meg a Magyar Számítógépes Nyelvészeti Konferenciát. Örömmre szolgál, hogy a rendezvény – az előző évek hagyományaihoz hasonlóan – fokozott érdeklődést váltott ki az ország nyelv- és beszédtechnológiai szakembereinek körében. A konferencia fő célja továbbra is a nyelv- és beszédtechnológia területén végzett legújabb, illetve folyamatban levő kutatási eredményeinek ismertetése és megvitatása, továbbá az esemény lehetőséget biztosít különféle hallgatói projektek, illetve ipari alkalmazások bemutatására is.

Idén a konferenciafelhívásra szép számban beérkezett tudományos előadások közül a programbizottság 46-ot fogadott el, így 32 előadás és 14 poszter-, illetve laptopos bemutató gazdagítja a konferencia programját.

Nagy örömet jelent számomra az is, hogy az idei konferencián – külön szekció keretében – kiemelt figyelmet szentelünk a szemantikus keresés terén elért eredményeknek. A számítógépes nyelvészet egyik legintenzívebben kutatott területéhez kapcsolódik a MASZEKER projekt, melynek keretében a Nemzeti Kutatási és Technológiai Hivatal is támogatja a rendezvényt. A projekt részleteiről több előadásból, poszterből és laptopos bemutatóból is informálódhat az érdeklődő közönség.

Az eddigi alkalmakhoz hasonlóan idén is tervezzük a „Legjobb Ifjú Kutatói Díj” odaítélését, mellyel a fiatal korosztály tagjait kívánjuk ösztönözni arra, hogy kiemelkedő eredményekkel járuljanak hozzá a magyarországi nyelv- és beszédtechnológiai kutatásokhoz. A díjat az MTA Számítástechnikai és Automatizálási Kutatóintézete ajánlotta fel az idei évben.

Szeretnék köszönetet mondani a programbizottságnak: Vámos Tibor programbizottsági elnöknek, valamint Alberti Gábor, Gordos Géza, László János, Prószéky Gábor és Váradi Tamás programbizottsági tagoknak. Szeretném továbbá megköszönni a rendezőbizottság (Alexin Zoltán, Almási Attila, Vincze Veronika) és a kötet szerkesztők (Tanács Attila, Vincze Veronika) munkáját is.

Csirik János, a rendezőbizottság elnöke

Szeged, 2010. november

III. Szemantika

Vonzatkeretlisták helyett polarításos hatáslánccsaládok – avagy a \mathfrak{ReALIS} σ függvénye	113
<i>Alberti Gábor, Kilián Imre</i>	
Személynév-egyértelműsítés a magyar weben	127
<i>Nagy T. István, Farkas Richárd</i>	
A Magyar WordNet felhasználhatósága lexikális jelentés-egyértelműsítésben.....	137
<i>Kuti Judit, Darja Fišer</i>	
A metaforikus nyelvhasználat korpuszalapú elemzése	145
<i>Babarczy Anna, Bencze Ildikó, Fekete István, Simon Eszter</i>	

IV. (Szemantikus) keresés

MASZEKER: projekt szemantikus keresőtechnológia kidolgozására	159
<i>Szőts Miklós, Csirik János, Gergely Tamás, Karvalics László</i>	
Nyelvészeti problémák a szabadalmak feldolgozásában.....	168
<i>Vincze Veronika, Nagy Ágoston, Klausz Ágnes, Almási Attila, Kiss Márton</i>	
Vonzatkeretek vizsgálata orvostudományi tárgyú, angol nyelvű szabadalmi szövegeken	180
<i>Klausz Ágnes, Vincze Veronika, Nagy Ágoston, Almási Attila</i>	
Egy vertikális nyelvi kereső készítése	190
<i>Orosz György</i>	

V. Beszédtechnológia

Környezetfüggetlen és sztochasztikus nyelvtanok összehasonlítása többnyelvű gépi beszéd felismerési feladatban	203
<i>Mozsolics Tamás, Tarján Balázs, Mihajlik Péter, Fegyó Tibor</i>	
Magyar nyelvű nagyszótáros beszéd felismerési feladatok adatelégtelenségi problémáinak csökkentése nyelvmodell-interpoláció alkalmazásával.....	216
<i>Tarján Balázs, Mihajlik Péter</i>	
Kulcsszókeresési kísérletek hangzó híryanagyokon beszédhang alapú felismerési technikákkal.....	224
<i>Gosztolya Gábor, Tóth László</i>	
Szótagok automatikus osztályozása spontán beszédben spektrális és prozódiai jellemzők alapján	236
<i>Beke András, Szaszák György</i>	

Bizonytalanságot jelölő kifejezések és hatókörük azonosítása természetes nyelvi szövegekben: a CoNLL-2010 verseny tapasztalatai	354
<i>Farkas Richárd, Vincze Veronika, Móra György, Csirik János, Szarvas György</i>	
Szemantikus annotációk létrehozása a weben nyelvtechnológiai eszközök támogatásával.....	358
<i>Héder Mihály</i>	
Melléknevek szűk szemantikai osztályainak detekciója a Magyar Nemzeti Szövegtárban jelentés-egyértelműsítés céljából	360
<i>Héja Enikő, Takács Dávid</i>	
Egy nyelvészeti UIMA-folyamat a kézi annotálástól az eredmények megjelenítéséig.....	362
<i>Kiss Márton, Nagy Ágoston</i>	
A MASZEKER felhasználói felületének kialakítása	365
<i>Minkó Mihály</i>	
Bűnügyi névelem-felismerés	366
<i>Molnár Gábor József, Kojedzinszky Tamás, Farkas Richárd</i>	
Igei igenevek problémája számítógépes nyelvészeti szempontból	371
<i>Nádasdi Péter</i>	
Terminológiai kivonatolás francia nyelvű szabadalmak leírásaiból különböző módszerek segítségével	375
<i>Nagy Ágoston</i>	
Szótáralapú kémiai NE-felismerő rendszer.....	379
<i>Nyilas Sándor, Németh Gábor, Almási Attila</i>	
Lényegkiemelő módszerek összehasonlítása közlekedési zajban történő beszéd felismerés céljából.....	384
<i>Sárosi Gellért, Tobler Zoltán, Mihajlik Péter, Fegyő Tibor</i>	
Valós idejű szövegosztályozás a Wikipedia szolgálatában.....	389
<i>Solt Illés, Héder Mihály, Tikk Domonkos</i>	
A HG-1 treebank: a nyelvtanírástól az online konkordanciáig	391
<i>Tóth Ágoston</i>	
Szerzői index, névmutató.....	395

Tartalomjegyzék

I. Információkinyerés

Panaszlevelek szerkezetének gépi felismerése.....	3
<i>Bártházi Eszter, Héder Mihály</i>	
OpinHu: online szövegek többnyelvű véleményelemzése	14
<i>Miháltz Márton</i>	
Videókhoz kapcsolódó kiegészítő információk többnyelvű keresése a Wikipédia segítségével	24
<i>Gyarmati Ágnes, Gareth J.F. Jones</i>	
DBPedia magyar nyelvű szövegek elemzéséhez.....	26
<i>Németh Bottyán, Vándor Tamás</i>	
Kontextualizált névelem-felismerés és relációkinyerés kórházi zárójelentésekben ..	35
<i>Solt Illés, Szidarovszky P. Ferenc, Tikk Domonkos</i>	
Kulcsszókinyerés magyar nyelvű tudományos publikációkból	47
<i>Berend Gábor, Farkas Richárd</i>	
Bibliográfiai hivatkozások automatikus kinyerése.....	56
<i>Váradai Tamás, Pintér Tibor, Mittelholcz Iván, Peredy Márta</i>	

II. Párhuzamos korpuszok

Statisztikai és hibrid módszerek párhuzamos korpuszok feldolgozására	69
<i>Laki László János, Prószéky Gábor</i>	
Többszavas kifejezések kezelése a párhuzamos korpuszokra épülő szótárkészítési módszertanban.....	80
<i>Héja Enikő, Sass Bálint</i>	
Félig kompozicionális szerkezetek a SzegedParalell angol–magyar párhuzamos korpuszban	91
<i>Vincze Veronika, Felvégi Zsuzsanna, R. Tóth Krisztina</i>	
Párhuzamos igei szerkezetek közvetlen kinyerése párhuzamos korpuszból.....	102
<i>Sass Bálint</i>	

Spontán beszédben rejlő nem verbális hangjelenségek – érzelmek, hanggesztusok – vizsgálata.....	249
<i>Vicsi Klára, Sztahó Dávid, Kiss Gábor, Czira Anita</i>	
Érzelmek automatikus osztályozása spontán beszédben.....	261
<i>Sztahó Dávid, Imre Viktor, Vicsi Klára</i>	

VI. Morfológia, korpusz

Ismeretlen kifejezések és a szófaji egyértelműsítés	275
<i>Zsibrita János, Vincze Veronika, Farkas Richárd</i>	
Obi-ugor morfológiai elemzők és korpuszok	284
<i>Fejes László, Novák Attila</i>	
A magyar frazeológiai adatbázis létrehozása és az ebből generált szinonim frazémaszótár munkálatai	292
<i>Bárdosi Vilmos, Kiss Gábor</i>	
Nyelvtechnológiai módszerek a Budapesti Szociolingvisztikai Interjú lexikai és szintaktikai vizsgálatában	300
<i>Várad Tamás, Peredy Márta, Oravecz Csaba</i>	

VII. Gépi tanulás

Szótáralapú névelem-felismerés szóhatárainak javítása gépi tanulási módszerrel..	317
<i>Móra György, Farkas Richárd</i>	
Klaszterek helyett prototípusok	325
<i>Kálmán László, Rung András</i>	
Főnévi csoportok azonosítása szabályalapú és hibrid módszerekkel.....	333
<i>Recski Gábor</i>	

VIII. Poszterek, laptopos bemutatók

Online morfológiai elemzők és szóalak-generátorok kisebb uráli nyelvekhez	345
<i>Bakró-Nagy Marianne, Endrédi István, Fejes László, Novák Attila, Oszkó Beatrix, Prószyék Gábor, Szeverényi Sándor, Várnai Zsuzsa, Wagner-Nagy Beáta</i>	
MSD-KR harmonizáció a Szeged Treebank 2.5-ben.....	349
<i>Farkas Richárd, Szeredi Dániel, Varga Dániel, Vincze Veronika</i>	