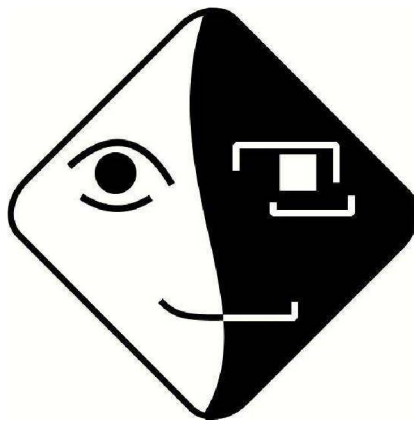


IX. Magyar Számítógépes Nyelvészeti Konferencia



MSZNY 2013

Szerkesztette:

Tanács Attila
Vincze Veronika

Szeged, 2013. január 7-8.
<http://www.inf.u-szeged.hu/mszny2013>

ISBN 978-963-306-189-3

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

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

Nyomtatta: JATEPress
6722 Szeged, Petőfi Sándor sugárút 30–34.

Szeged, 2012. december

Előszó

2013. január 7-8-án kilencedik alkalommal rendezzük meg Szegeden a Magyar Számítógépes Nyelvészeti Konferenciát. A konferencia fő célja – a hagyományokhoz híven – a nyelv- és beszédtechnológia területén végzett legújabb, illetve folyamatban levő kutatások eredményeinek ismertetése és megvitatása, mindemellett lehetőség nyílik különféle hallgatói projektek, illetve ipari alkalmazások bemutatására is. A korábbi évekhez hasonlóan, a rendezvény fokozott érdeklődést váltott ki az ország nyelv- és beszédtechnológiai szakembereinek körében.

A konferenciafelhívásra szép számban beérkezett tudományos előadások közül a programbizottság 42-t fogadott el az idei évben, így 26 előadás és 16 poszter-, illetve laptopos bemutató gazdagítja a konferencia programját. A programban a magyar számítógépes nyelvészet rendkívül széles skálájáról találhatunk előadásokat a beszédtechnológiától kezdve a számítógépes morfológia és szintaxis területén át az információkinyerésig és gépi fordításig.

Nagy örömet jelent számomra az is, hogy Gósy Mária, a Nyelvtudományi Intézet Fonetikai Osztályának tudományos osztályvezetője, az ELTE BTK Fonetika Tanszékének tanszékvezető egyetemi tanára elfogadta meghívásunkat, és *Spontán beszéd: szabályok és szabálytalanságok* című plenáris előadása is a konferenciaprogram részét képezi.

Ahogy az már hagyománnyá vált, 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 díj felajánlásáért az MTA Számítástechnikai és Automatizálási Kutatóintézetének tartozunk köszönettel.

Szeretnék köszönetet mondani a programbizottságnak: Vámos Tibor programbizottsági elnöknek, valamint Alberti Gábor, Gordos Géza, Kornai András, László János, Prószéky Gábor és Váradi Tamás programbizottsági tagoknak. Szeretném továbbá megköszönni a rendezőbizottság és a kötet szerkesztők munkáját is.

Csirik János, a rendezőbizottság elnöke

Szeged, 2012. december

Tartalomjegyzék

I. Beszédtechnológia, fonológia

Mély neuronhálók az akusztikus modellezésben	3
<i>Grósz Tamás, Tóth László</i>	
Magyar nyelvű, kísérleti e-mail diktáló rendszer	13
<i>Tarján Balázs, Nagy Tímea, Mihajlik Péter, Fegyó Tibor</i>	
Hogyan tanuljunk kevés információból is? A RIP algoritmus továbbfejlesztett változatai	21
<i>Biró Tamás</i>	

II. Lexikológia, fordítás

Angol nyelvű összetett főnevek értelmezése parafrázisok segítségével	35
<i>Dobó András, Stephen G. Pulman</i>	
Félig kompozicionális szerkezetek automatikus felismerése doménadaptációs technikák segítségével a Szeged Korpuszon	47
<i>Nagy T. István, Vincze Veronika, Zsibrita János</i>	
Automatikusan generált online szótárak: az EFNILEX projekt eredményei	59
<i>Héja Enikő, Takács Dávid</i>	
A 4lang fogalmi szótár	62
<i>Kornai András, Makrai Márton</i>	
Hunglish mondattan – átrendezésalapú angol–magyar statisztikai gépfordító-rendszer	71
<i>Laki László János, Novák Attila, Siklósi Borbála</i>	

III. Korpusznyelvészet

Nyelvtanfejlesztés, implementálás és korpuszépítés: A HunGram 2.0 és a HG-1 Treebank legfontosabb jellemzői	85
<i>Laczkó Tibor, Rákosi György, Tóth Ágoston, Csernyi Gábor</i>	
HunLearner: a magyar nyelv nyelvtanulói korpusza.....	97
<i>Vincze Veronika, Zsibrita János, Durst Péter, Szabó Martina Katalin</i>	
Automatikus korpuszépítés tulajdonnév-felismerés céljára	106
<i>Nemeskey Dávid Márk, Simon Eszter</i>	

IV. Pszichológia

- Szemantikus szerepek a narratív kategoriális elemzés (NARRCAT) rendszerében ... 121
Ehmann Bea, Lendvai Piroska, Miháلتz Márton, Vincze Orsolya, László János
- A Regresszív Képzelti Szótár magyar nyelvű változatának létrehozása 124
Pólya Tibor, Szász Levente

V. Morfológia, szintaxis

- Helyesírás.hu – Nyelvtechnológiai megoldások automatikus helyesírási tanácsadó rendszerben 135
Miháلتz Márton, Hussami Péter, Ludányi Zsófia, Mittelholcz Iván, Nagy Ágoston, Oravecz Csaba, Pintér Tibor, Takács Dávid
- Helyesírási hibák automatikus javítása orvosi szövegekben a szövegkörnyezet figyelembevételével..... 148
Siklósi Borbála, Novák Attila, Prószéky Gábor
- Magyar nyelvű klinikai rekordok morfológiai egyértelműsítése 159
Orosz György, Novák Attila, Prószéky Gábor
- O & közèpmağar zoalactanğ èlèmzo, 170
Novák Attila, Wenszky Nóra
- Domének közti hasonlóságok és különbségek a szófajok és szintaktikai viszonyok eloszlásában 182
Vincze Veronika
- Gondolatok a (magyar) statisztikai szintaktikai elemzőkről 193
Farkas Richárd

VI. Szemantika

- A lehetőségalmazok meghatározása az inkvizitív szemantikában..... 205
Szécsényi Tibor
- Magyar és angol szavak szemantikai hasonlóságának automatikus kiszámítása 213
Dobó András, Csirik János
- A ReALIS tudástároló és következtető alrendszere 225
Kilián Imre
- Az igazság pillanata – avagy a ReALIS α horgonyzó függvénye 236
Alberti Gábor, Károly Márton, Kilián Imre, Kleiber Judit, Vadász Noémi

VII. Információkinyerés és -visszakeresés

Kulcsszókinyerés alapú dokumentumklaszterezés	251
<i>Berend Gábor, Farkas Richárd, Vincze Veronika, Zsibrita János, Jelasity Márk</i>	
Információorientált dokumentumosztályozás a magyar Wikipédián.....	263
<i>Subecz Zoltán, Farkas Richárd</i>	
Frame-szemantikára alapozott információ-visszakereső rendszer	275
<i>Szóts Miklós, Gyarmathy Zsófia, Simonyi András</i>	

VIII. Posztterek és laptopos bemutatók

Dokumentumcsoportok automatikus kulcsszavazása és témakövetés	289
<i>Ács Zsombor, Farkas Richárd</i>	
Egy hatékonyabb webes sablonszűrő algoritmus –avagy miként lehet a cumisüveg potenciális veszélyforrás Obamára nézve	297
<i>Endrédi István, Novák Attila</i>	
MASZEKER: szemantikus kereső program	302
<i>Hussami Péter</i>	
PureToken: egy új tokenizáló eszköz	305
<i>Indig Balázs</i>	
Ismeretlen szavak helyes kezelése kötegelt helyesírás-ellenőrző programmal	310
<i>Indig Balázs, Prószték Gábor</i>	
A ReALIS statikus interpretációjának kísérleti implementációja	318
<i>Károly Márton</i>	
A szövegtörzsek szókincsének összehasonlítása szótári címszójegyzék felhasználásával – neologizmusok és archaizmusok detektálása	324
<i>Kiss Gábor, Kiss Márton</i>	
Morfológiai egyértelműsítés nyelvfüggetlen annotáló módszerek kombinálásával	331
<i>Laki László János, Orosz György</i>	
Anonimizálási gyakorlat? – Egy magyar korpusz anonimizálásának tanulságai	338
<i>Mátyus Kinga</i>	
OpinHuBank: szabadon hozzáférhető annotált korpusz magyar nyelvű véleményelemzéshez.....	343
<i>Miháltz Márton</i>	
Miből lesz a robot MÁV-pénztáros?	346
<i>Nemeskey Dávid, Recski Gábor, Zséder Attila</i>	

Az új magyar Braille-rövidírás korpuszvezérelt kialakításának lehetőségei.....	348
<i>Sass Bálint</i>	
Neticle – Megmutatjuk, mit gondol a web	351
<i>Szekeres Péter</i>	
Vektortér alapú szemantikai szóhasonlósági vizsgálatok	354
<i>Tóth Ágoston</i>	
Magyar nyelvű néprajzi keresőrendszer.....	361
<i>Zsibrita János, Vincze Veronika</i>	
magyarlanc 2.0: szintaktikai elemzés és felgyorsított szófaji egyértelműsítés.....	368
<i>Zsibrita János, Vincze Veronika, Farkas Richárd</i>	
Szerzői index, névmutató.....	375