

XII. Magyar Számítógépes Nyelvészeti Konferencia



MSZNY 2016

Szerkesztette:

Tanács Attila
Varga Viktor
Vincze Veronika

Szeged, 2016. január 21-22.
<http://rgai.inf.u-szeged.hu/mszny2016>

ISBN: 978-963-306-450-4

Szerkesztette: Tanács Attila, Varga Viktor és Vincze Veronika
{tanacs, vincze}@inf.u-szeged.hu
viktor.varga.1991@gmail.com

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

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

Szeged, 2016. január

Előszó

2016. január 21-22-én immár tizenkettedik alkalommal rendezzük meg Szegeden a Magyar Számítógépes Nyelvészeti Konferenciát. A konferencia fő célkitűzése a kezdetek óta mit sem változott: a rendezvény fő célja 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.

Örömmre szolgál, hogy a hagyományokat követve a konferencia idén is nagyfokú érdeklődést váltott ki az ország nyelv- és beszédtechnológiai szakembereinek körében. A konferenciafelhívásra idén is nagy számban beérkezett tudományos cikkek közül a programbizottság 25 előadást, 8 poszter-, illetve 4 laptopos bemutatót fogadott el. Újdonságot jelent, hogy egyes témákat mind az előadások, mind pedig a laptopos bemutatók között is megtalálunk, ezzel is lehetőséget adva a kutatási témák minél szélesebb körű bemutatására. A programban a magyar számítógépes nyelvészet rendkívül széles skálájáról találhatunk előadásokat a számítógépes morfológiától kezdve a beszédtechnológián át a szaknyelvi szövegek számítógépes feldolgozásáig. Mindemellett a magyar nyelvtechnológiai műhelyek együttműködésében megvalósuló, egy egységes magyar előfeldolgozó lánc kifejlesztését célzó INFRA projektnek is külön szekciót szentelünk.

Nagy örömet jelent számomra az is, hogy Pléh Csaba, az MTA rendes tagja elfogadta meghívásunkat, és plenáris előadása is gyarapítja a konferencia résztvevőinek szakmai ismereteit.

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 konferencia sikeréhez a Neumann János Számítógép-tudományi Társaság szíves anyagi támogatása is hozzájárul, illetőleg a konferencia fogadása a MeltWater R&D nagylelkű támogatásával valósul meg. A rendezőbizottság nevében ezúton is szeretném kifejezni hálás köszönetünket mindkét támogatóknak.

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

Csirik János, a rendezőbizottság elnöke

Szeged, 2016. január

Tartalomjegyzék

I. Fordítás

Beágyazási modellek alkalmazása lexikai kategorizációs feladatokra	3
<i>Siklósi Borbála, Novák Attila</i>	
Building Definition Graphs using Monolingual Dictionaries of Hungarian .	15
<i>Gábor Recski, Attila Bolevác, Gábor Borbély</i>	
Közeli rokonunk, az autó	27
<i>Siklósi Borbála, Novák Attila</i>	
Gépi fordítás minőségbecslésének optimalizálása kétnyelvű szótár és WordNet segítségével	37
<i>Yang Zijian Győző, Laki László</i>	

II. Morfológia, előfeldolgozás

Ékezetek automatikus helyreállítása magyar nyelvű szövegekben	49
<i>Novák Attila, Siklósi Borbála</i>	
Utilizing Word Embeddings for Part-of-Speech Tagging	59
<i>Gábor Berend</i>	
Módosított morfológiai egyértelműsítés és integrált konstituenselemzés a magyarlanc 3.0-ban	68
<i>Farkas Richárd, Szántó Zsolt, Vincze Veronika, Zsibrita János</i>	
Új integrált magyar morfológiai elemző	78
<i>Novák Attila</i>	

III. Beszédtechnológia

Magyar nyelvű, élő közéleti- és hírműsorok gépi feliratozása	89
<i>Tarján Balázs, Varga Ádám, Tobler Zoltán, Szaszák György, Fegyő Tibor, Bordás Csaba, Mihajlik Péter</i>	
Egy magyar nyelvű beszédfelismerő rendszer szószintű hibáinak elemzése .	100
<i>Gosztolya Gábor, Vincze Veronika, Grósz Tamás, Tóth László</i>	
Szövegalapú nyelvi elemző kiértékelése gépi beszédfelismerő hibákkal terhelt kimenetén	111
<i>Tündik Máté Ákos, Szaszák György</i>	
Nevetések automatikus felismerése mély neurális hálók használatával	122
<i>Gosztolya Gábor, Beke András, Neuberger Tilda</i>	

Magyar nyelvű szövegek automatikus fonetikai átírása	134
<i>Novák Attila, Siklósi Borbála</i>	

Gépi beszéd természetességének növelése automatikus, beszédjel alapú hangsúlycímkező algoritmussal	144
<i>Szaszák György, Beke András, Olaszy Gábor, Tóth Bálint Pál</i>	

Mély neuronhálós akusztikus modellek gyors adaptációja multi-taszka tanítással	154
<i>Tóth László, Gosztolya Gábor</i>	

IV. Szemantika, szentimentelemzés

Angol és magyar nyelvű kérdések a számítógépes nyelvészetben	165
<i>Vincze Veronika</i>	

Aspektusszintű annotáció és szentimentet módosító elemek egy magyar nyelvű szentimentkorpuszban	174
<i>Szabó Martina Katalin, Vincze Veronika, Hangya Viktor</i>	

Az érzelmek beszédre gyakorolt hatása, azaz a spontán beszéd szintaxisának érzelmekkel való kapcsolata a HuComTech Korpuszban	183
<i>Kiss Hermína</i>	

Rádióműsorok elemzése a WordNetAffect érzelmi szótár segítségével	193
<i>Lukács Gergely, Martos Tamás, Jani Máttyás, Takács György</i>	

V. Szaknyelv, speciális nyelvhasználat

A magyar jelnyelvi korpusz létrehozásának és annotálásának kihívásai . . .	207
<i>Bartha Csilla, Varjasi Szabolcs, Holecz Margit</i>	

Jogszabályok hivatkozásainak automatikus felismerése és a belső hivatkozások struktúrája	220
<i>Hamp Gábor, Syi, Markovich Réka</i>	

Digitális Konzílium – egy szemészeti klinikai keresőrendszer	230
<i>Siklósi Borbála, Novák Attila</i>	

VI. Szintaxis

Egyszer „van”, hol nem „van”: A létige kezelése függőségi nyelvtanokban . .	243
<i>Simkó Katalin Ilona, Vincze Veronika</i>	

Szabályalapú szintaktikai elemző szintaktikai szabályok nélkül	251
<i>Kovács Viktória, Simkó Katalin Ilona, Szécsényi Tibor</i>	

Mozaik nyelvmodell az ANAGRAMMA elemzőhöz	260
<i>Indig Balázs, Laki László, Prószék Gábor</i>	

VII. Poszterek

Discovering Utterance Fragment Boundaries in Small Unsegmented Texts <i>László Drienkó</i>	273
Magyar nyelvű orvosi szakcikkek hivatkozásainak automatikus feldolgozása <i>Farkas Richárd, Kojedzinszky Tamás, Sliz-Nagy Alex, Tímár György, Zsibrita János</i>	282
Többsávós, zajtűrő beszédfelismerés mély neuronhálóval <i>Kovács György, Tóth László</i>	287
Statistikai koreferenciafeloldó rendszer magyar nyelvre — első eredmények <i>Munkácsy Gergely, Farkas Richárd</i>	295
Angol-magyar többszavas kifejezések szótárának automatikus építése párhuzamos korpuszok segítségével <i>Nagy T. István, Vincze Veronika</i>	298
A magabiztosság-krízis skála alkalmazása idegen nyelvű megnyilatkozásoknál <i>Puskás László, Pólya Tibor</i>	305
A magyar Wikipédia automatikus bejárása és elemzése <i>Simkó Marcell, Góth Júlia</i>	313
Univerzális dependencia és morfológia magyar nyelvre <i>Vincze Veronika, Farkas Richárd, Simkó Katalin Ilona, Szántó Zsolt, Varga Viktor</i>	322

VIII. Laptopos bemutatók

Lórum ipse: magyar vakszöveg-generátor <i>Nagy Viktor, Takács Dávid</i>	333
--	-----

IX. Angol nyelvű absztraktok

Van's upon a Time: Copulas in Dependency Grammars <i>Katalin Ilona Simkó, Veronika Vincze</i>	337
Névmutató	339

