

XI. Magyar Számítógépes Nyelvészeti Konferencia



MSZNY 2015

Szerkesztette:

Tanács Attila
Varga Viktor
Vincze Veronika

Szeged, 2015. január 15-16.
<http://rgai.inf.u-szeged.hu/mszny2015>

ISBN: 978-963-306-359-0

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

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

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

Szeged, 2015. január

Előszó

Idén immár tizenegyedik alkalommal rendezzük meg Szegeden a Magyar Számítógépes Nyelvészeti Konferenciát 2015. január 15-16-án. A konferencia fő célkitűzése a kezdetek óta állandó maradt: a rendezvény fő profilja 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.

Nagy örömmre szolgál, hogy a hagyományoknak megfelelően a konferencia nagyfokú érdeklődést váltott ki az ország nyelv- és beszédtechnológiai szakembereinek körében. A konferenciafelhívásra idén is nagy számban beérkezett tudományos előadások közül a programbizottság 36-ot fogadott el az idei évben, így 24 előadás, 8 poszter-, illetve 4 laptopos bemutató gazdagítja a konferencia programját. A programban a magyar számítógépes nyelvészet rendkívül széles skálájáról található előadásokat a számítógépes szintaxis és szemantika területétől kezdve a véleménykinyerésen át a klinikai szövegek számítógépes feldolgozásáig.

Nagy örömet jelent számomra az is, hogy Tihanyi László, az Európai Bizottság gépi fordítással foglalkozó szakértője, elfogadta meghívásunkat, és plenáris előadása is a konferenciaprogram szerves 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.

Ezúton szeretném megköszönni a Neumann János Számítógép-tudományi Társaságnak szíves anyagi támogatásukat.

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

Csirik János, a rendezőbizottság elnöke

Szeged, 2015. január

Tartalomjegyzék

I. Fordítás

Gépi fordítás minőségének becslése referencia nélküli módszerrel	3
<i>Yang Zijian Győző, Laki László, Prószéky Gábor</i>	
Synonym Acquisition from Translation Graph	14
<i>Judit Ács</i>	
Comparison of Distributed Language Models on Medium-resourced Languages	22
<i>Márton Makrai</i>	
Statisztika megbízhatósága a nyelvészetben – Széljegyzetek egy szótárbővítés ürügyén	34
<i>Naszódi Mátyás</i>	

II. Szintaxis, szemantika

Konstituensfák automatikus átalakítása függőségi fákká vagy kézi annotáció?	49
<i>Simkó Katalin Ilona, Vincze Veronika, Szántó Zsolt, Farkas Richárd</i>	
Hungarian Data-Driven Syntactic Parsing in 2014	61
<i>Zsolt Szántó, Richárd Farkas, Anders Björkelund, Özlem Çetinoğlu, Agnieszka Faleńska, Thomas Müller, Wolfgang Seeker</i>	
Nyelvadaptáció a többszavas kifejezések automatikus azonosításában	71
<i>Nagy T. István, Vincze Veronika</i>	
Lexikális behelyettesítés magyarul	83
<i>Takács Dávid, Gábor Kata</i>	
Szemantikus szerepek automatikus címkézése függőségi elemző alkalmazásával magyar nyelvű gazdasági szövegeken	95
<i>Subecz Zoltán</i>	

III. Morfológia, korpusz

Mennyiségből minőséget: Nyelvtechnológiai kihívások és tanulságok az MNSz új változatának elkészítésében	109
<i>Oravecz Csaba, Sass Bálint, Váradí Tamás</i>	

Magyar nyelvű webes szövegek morfológiai és szintaktikai annotációja ...	122
<i>Vincze Veronika, Varga Viktor, Papp Petra Anna, Simkó Katalin Ilona, Zsibrita János, Farkas Richárd</i>	

Finnugor nyelvű közösségek nyelvtechnológiai támogatása online tartalmak létrehozásában	133
<i>Benyeda Ivett, Koczka Péter, Ludányi Zsófia, Simon Eszter, Váradi Tamás</i>	

„Olcsó” morfológia	145
<i>Novák Attila</i>	

IV. Beszédtechnológia

Kétszintű algoritmus spontán beszéd prozódiaalapú szegmentálására	161
<i>Beke András, Markó Alexandra, Szaszák György, Váradi Viola</i>	

Környezetfüggő akusztikai modellek létrehozása Kullback-Leibler– divergencia alapú klaszterezéssel	174
<i>Grósz Tamás, Gosztolya Gábor, Tóth László</i>	

Hibajavítási idő csökkentése magyar nyelvű diktálórendszerben	182
<i>Szabó Lili, Tarján Balázs, Mihajlik Péter, Fegyő Tibor</i>	

V. Véleménykinyerés

TrendMiner: politikai témájú Facebook-üzenetek feldolgozása és szociálpszichológiai elemzése	195
<i>Miháltz Márton, Váradi Tamás</i>	

A véleményváltozás azonosítása politikai témájú közösségi médiában megjelenő szövegekben	198
<i>Pólya Tibor, Csertő István, Fülöp Éva, Kóvágó Pál, Miháltz Márton, Váradi Tamás</i>	

Doménspecifikus polarításlexikonok automatikus előállítására magyar nyelvre	210
<i>Hangya Viktor, Farkas Richárd</i>	

Egy magyar nyelvű szentimentkorpusz létrehozásának tapasztalatai	219
<i>Szabó Martina Katalin, Vincze Veronika</i>	

Entitásorientált véleménydetekció webes híryanagokból	227
<i>Hangya Viktor, Farkas Richárd, Berend Gábor</i>	

VI. Alkalmazások

Nem felügyelt módszerek alkalmazása releváns kifejezések azonosítására és csoportosítására klinikai dokumentumokban	237
<i>Siklósi Borbála, Novák Attila</i>	

Tartalomjegyzék	VII
-----------------	-----

Az enyhe kognitív zavar automatikus azonosítása beszédátiratok alapján .	249
<i>Vincze Veronika, Hoffmann Ildikó, Szatlóczki Gréta, Bíró Edit, Gosztolya Gábor, Tóth László, Pákáski Magdolna, Kálmán János</i>	

Beszéd-zene lejátszási listák nyelvtechnológiai vonatkozása	257
<i>Benyeda Ivett, Jani Máttyás, Lukács Gergely</i>	

VII. Poszterbemutatók

Gyógyszermellékhatások kinyerése magyar nyelvű orvosi szaklapok szövegeiből	271
<i>Farkas Richárd, Miklós István, Tímár György, Zsibrita János</i>	

Elliptikus listák jogszabálysövegekben	273
<i>Hamp Gábor, Syi, Markovich Réka</i>	

FinUgRevita: nyelvtechnológiai eszközök fejlesztése kisebbségi finnugor nyelvekre	282
<i>Horváth Csilla, Kozmács István, Szilágyi Norbert, Vincze Veronika, Nagy Ágoston, Bogár Edit, Fenyvesi Anna</i>	

Az automatikus irreguláriszöngge-detekció sikeressége az irregularitás mintázatának függvényében magyar (spontán és olvasott) beszédben	290
<i>Markó Alexandra, Csapó Tamás Gábor</i>	

Igei vonzatkeretek és tematikus szerepek felismerése nyelvi erőforrások összekapcsolásával egy kereslet-kínálat elvű szövegelemzőben	298
<i>Miháltz Márton, Indig Balázs, Prószéky Gábor</i>	

28 millió szintaktikailag elemzett mondat és 500000 igei szerkezet	303
<i>Sass Bálint</i>	

Egy kereslet-kínálat elvű elemző működése és a koordináció kezelésének módszere	309
<i>Sass Bálint</i>	

SzegedKoref: kézzel annotált magyar nyelvű koreferenciakörpusz	312
<i>Vincze Veronika, Hegedűs Klára, Farkas Richárd</i>	

VIII. Laptopos bemutatók

Yako: egy intelligens üzenetváltó alkalmazás nyelvtechnológiai kihívásai . .	323
<i>Farkas Richárd, Kojedzinszky Tamás, Zsibrita János, Wieszner Vilmos</i>	

HumInA projektcsoport a ReALIS1.1 bázisán	326
<i>Nóthig László, Alberti Gábor</i>	

Neticle – Megmutatjuk, mit gondol a web	333
<i>Szekeres Péter</i>	

Magyar nyelvű hasonló tartalmú orvosi leletek azonosítása	336
<i>Wieszner Vilmos, Farkas Richárd, Csizmadia Sándor, Palkó András</i>	

IX. Angol nyelvű absztraktok

Natural Language Processing for Mixed Speech-Music Playlist Generation	341
<i>Ivett Benyeda, Mátyás Jani, Gergely Lukács</i>	

The Reliability of Statistics in Linguistics Notes to a Dictionary Extension	342
<i>Mátyás Naszódi</i>	

Automatic Conversion of Constituency Trees into Dependency Trees or Manual Annotation?	344
<i>Katalin Ilona Simkó, Veronika Vincze, Zsolt Szántó, Richárd Farkas</i>	

SzegedKoref: A Manually Annotated Coreference Corpus of Hungarian . .	345
<i>Veronika Vincze, Klára Hegedűs, Richárd Farkas</i>	

Morphological and Syntactic Annotation of Hungarian Webtext	346
<i>Veronika Vincze, Viktor Varga, Petra Anna Papp, Katalin Ilona Simkó, János Zsibrita, Richárd Farkas</i>	

Névmutató	347
---------------------	-----