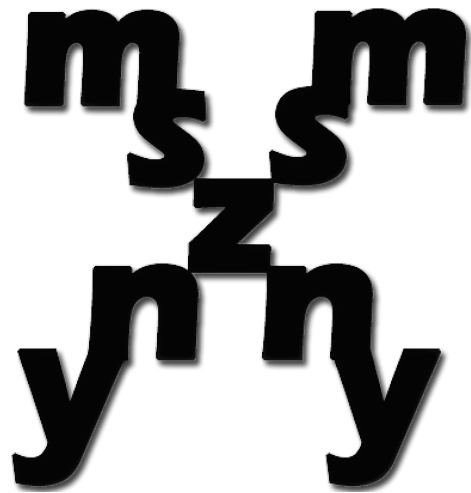


XIX. Magyar Számítógépes Nyelvészeti Konferencia



Szerkesztette:
Berend Gábor
Gosztolya Gábor
Vincze Veronika

Szeged, 2023. január 26–27.

Szerkesztette:

Berend Gábor, Gosztolya Gábor, Vincze Veronika
{berendg,ggabor,vinczev}@inf.u-szeged.hu

Felelős kiadó:

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

ISBN: 978-963-306-912-7

Nyomtatta:

Innovariant Nyomdaipari Kft.
6750 Algyő, Ipartelep 4.

Szeged, 2023. január

Az MSZNY 2023 konferencia szervezője:

ELKH-SZTE Mesterséges Intelligencia Kutatócsoport

Előszó

2023. január 26–27-én már tizenkilencedik alkalommal kerül sor a Magyar Számítógépes Nyelvészeti Konferencia megrendezésére. Ebben az évben először hibrid formátumban kerül lebonyolításra a konferencia, lehetőséget teremtve mind a személyes részvételre, mind pedig a konferencia élő közvetítésének nyomon követésére.

A konferencia fő célkitűzése a kezdetek óta állandó: lehetőséget biztosítani a nyelv- és beszédtechnológia területén végzett kutatások eredményeinek ismertetésére és megvitatására, ezen felül a különféle hallgatói projektek, illetve ipari alkalmazások bemutatására. A hagyományokat követve a konferencia idén is nagyfokú érdeklődést váltott ki az ország nyelv- és beszédtechnológiai szakembereinek körében. A 43 beküldött cikkből gondos mérlegelést követően 35 cikk került elfogadásra, melyek témája számos szakterületre terjed ki a legújabb nyelvi modellek bemutatásától kezdve a beszédtechnológia eredményein keresztül a gépi fordításig.

Nagy örömet jelent számunkra, hogy Tikk Domonkos elfogadta meghívásunkat, aki plenáris előadását *A Netflix Prizetól a Tabooláig: a Gravity és az ajánlórendszerek fejlődése* címmel fogja megtartani.

Az idei évben is különdíjjal jutalmazzuk a konferencia legjobb cikkét, mely a legjelentősebb eredményekkel járul hozzá a magyarországi nyelv- és beszédtechnológiai kutatásokhoz. Ezen felül immár ötödik alkalommal osztjuk ki a legjobb bíráló díját, amellyel a bírálók fáradtságos, ugyanakkor nélkülözhetetlen munkáját kívánjuk elismerni.

Köszönettel tartozunk továbbá a Neumann János Számítógéptudományi Társaság Mesterséges Intelligencia Szakosztályának a konferencia sikeres lebonyolításához nyújtott anyagi hozzájárulásáért.

A szervezőbizottság nevében,
Ács Judit, Berend Gábor, Gosztolya Gábor, Nemeskey Dávid Márk, Novák Attila,
Simon Eszter, Sztahó Dávid, Vincze Veronika

Tartalomjegyzék

Orvosi nyelv- és beszédtechnológia	1
3 Klinikai leletek strukturálása mondatszintű címkézéssel <i>Szabó Ledenyi Klaudia, Pusztai Ágnes, Kicsi András, Vidács László</i>	
17 Fokozás szkizofréniában <i>Szabó Martina Katalin, Vincze Veronika, Guba Csenge, Dam Bernadett, Solymos Adrienn, Bagi Anita, Szendi István</i>	
33 Sclerosis multiplex felismerése spontán beszédből wav2vec 2.0 modellekből kinyert jellemzőkkel <i>Gosztolya Gábor, José Vicente Egas-López, Svindt Veronika, Bóna Judit, Hoffmann Ildikó</i>	
45 A borderline személyiségzavar felismerése a nyelvhasználat lexikai és grammatikai jellemzői alapján <i>Felletár Fanni, Babarczy Anna</i>	
Szemantika	61
63 Magyar melléknevek poliszém jelentéseinek automatikus kinyerése gráfokkal <i>Héja Enikő, Ligeti-Nagy Noémi</i>	
77 Mondatszám-meghatározás hatása a magyar nyelvű jogi szövegek extraktív kivonatainak minőségére <i>Csányi Gergely Márk, Gadó Krisztián, Bajári Lúcia, Megyeri Andrea, Fülöp Anna, Egri Erika, Vági Renátó, Nagy Dániel, Vadász János Pál, Üveges István</i>	
91 Data Augmentation for Machine Translation via Dependency Subtree Swapping <i>Attila Nagy, Dorina Lakatos, Botond Barta, Patrick Nanyis, Judit Ács</i>	
107 Neurális entitásorientált szentimentelemző alkalmazás magyar nyelvre <i>Yang Zijian Győző, Laki László János</i>	
119 Koreferenciafeloldás magyar szövegeken BERT-tel <i>Vadász Noémi, Nyéki Bence</i>	
Beszédtechnológia	133
135 „Feeding the BEAST” – A BEA Speech Transcriber továbbfejlesztése és integrálása neurális nyelvmodellel <i>Kádár Máté Soma, Dobsinszki Gergely, Mády Katalin, Mihajlik Péter</i>	

- 145 Magyar nyelvű neurális beszédzintézis vizsgálata dialógus helyzetben
Zainkó Csaba, Csapó Tamás Gábor, Bartalis Mátyás, Németh Géza, Németh Norbert, Szász Gábor Krisztián, Szviridov István
- 159 Effects of emotional speech on forensic voice comparison using deep speaker embeddings
Mohammed Hamzah Abed, Dávid Sztahó
- 171 Cross-lingual dysphonic speech detection using pre-trained speaker embeddings
Aziz Dosti Ali Hama Salih, Dávid Sztahó

Korpuszok, adatbázisok

185

- 187 HunEmPoli: magyar nyelvű, részletesen annotált emóciókorpusz
Ring Orsolya, Vincze Veronika, Guba Csenge, Üveges István
- 203 MILQA kérdés-válasz benchmark adatbázis
Novák Attila, Novák Borbála
- 217 Hát te mekkorát nőttél! – A HuLU első életéve új adatbázisokkal és webszolgáltatással
Ligeti-Nagy Noémi, Héja Enikő, Laki László János, Takács Dávid, Yang Zijian Győző, Váradi Tamás
- 231 HunSum-1: an Abstractive Summarization Dataset for Hungarian
Botond Barta, Dorina Lakatos, Attila Nagy, Milán Konor Nyist, Judit Ács

Nyelvmodellek

245

- 247 Jönnnek a nagyok! BERT-Large, GPT-2 és GPT-3 nyelvmodellek magyar nyelvre
Yang Zijian Győző, Dodé Réka, Ferenczi Gergő, Héja Enikő, Jelencsik-Mátyus Kinga, Kőrös Ádám, Laki László János, Ligeti-Nagy Noémi, Vadász Noémi, Váradi Tamás
- 263 Látens szemantikus eloszlások használata a nyelvi modellek előtanítása során
Berend Gábor
- 275 Magyar nyelvű időjárásjelentések nyelvi modell alapú automatizált generálása
Knap Árpád, Dömsödi L Biborka, Mogyorósi Pálma, Szigeti Péter, Tóth Andor, Virág Attila, Kmetty Zoltán

Morfológia, előfeldolgozás

289

- 291 Korpusztisztítás és sorvégi kötőjelek kezelése karakteralapú neurális nyelvmodellel
Pethő Gergely, Sass Bálint, Simon László, Lipp Veronika
- 305 Transformer-alapú HuSpaCy előelemző láncok
Szabó Gergő, Orosz György, Szántó Zsolt, Berkecz Péter, Farkas Richárd
- 319 Hybrid lemmatization in HuSpaCy
Péter Berkecz, György Orosz, Zsolt Szántó, Gergő Szabó, Richárd Farkas
- 331 Neural Morphological Generators for Hungarian
László János Laki, Noémi Ligeti-Nagy, Noémi Vadász, Zijian Győző Yang
- 341 Gondolatok a gondola-tokról. Morfológiai annotációt javító módszerek tesztelése gold standard korpuszon
K. Molnár Emese, Dömötör Andrea

Poszter, laptopos bemutató

355

- 357 A beszéd artikulációs mozgásának predikciója agyi jel alapján – kezdeti eredmények
Csapó Tamás Gábor, Arthur Frigyes Viktor, Nagy Péter, Boncz Ádám
- 369 Magyarcentrikus többnyelvű gépfordító rendszerek létrehozása
Laki László János, Yang Zijian Győző
- 381 Többnyelvű modellek és PEGASUS finomhangolása magyar nyelvű absztraktív összefoglalás feladatára
Yang Zijian Győző
- 395 Insulinrezisztencia betegség jelenségének felismerése és osztályozása orvosi dokumentumokban
Yang Zijian Győző
- 405 huBERT alapú sziámi neurális háló architektúrák elemzése ügyfélszolgálati emailek klasszifikációjára
Vándor Péter, Csáki Csaba
- 417 HuBERTUSz: Alacsony paraméterszámú transzformer modellek létrehozása és kiértékelése magyar nyelvre
Ficsor Tamás, Berend Gábor
- 433 A new ParlaMint corpus for Hungarian – 30m tokens of annotated parliamentary data
Noémi Ligeti-Nagy, Réka Dodé, Kinga Jelencsik-Mátyus, Zsófia Varga, Enikő Héja, Tamás Váradi

- 447 Korpuszépítés és -feldolgozás leartott webes tartalomból
Kalcsó Gyula, Mihály Eszter, Szűcs Kata Ágnes
- 457 Tagmondatokra bontás és NP-chunking függőségi alapon
Dömötör Andrea, Nemeskey Dávid

Szerzői index, névmutató

471