

Lucrări științifice

Selectie: 2019-2025

- Radu Ion, Vasile Păiș, Verginica Barbu Mititelu, Elena Irimia, Valentin Badea, Dan Tufiș: Unsupervised Word Sense Induction using Transformer's Attention Mechanis. *Machine Learning and Knowledge Extraction*. 2025; 7(1):10.2025, <https://www.mdpi.com/2504-4990/7/1/10>
- Păiș, Vasile and Barbu Mititelu, Verginica and Irimia, Elena and Ion, Radu and Tufiș, Dan Under-Represented Speech Dataset from Open Data: Case Study on the Romanian Language in *Applied Sciences* 2024 14(19)
Special Issue on Natural Language Processing in the Era of Artificial Intelligence
<https://www.mdpi.com/2076-3417/14/19/9043>
- Vasile Păiș, Radu Ion, Elena Irimia, Verginica Barbu Mititelu, Valentin Badea, Dan Tufis, System for the anonymization of Romanian jurisprudence. In *Journal Artificial Intelligence and Law*, Springer, September 2024, ISSN 1572-8382, doi 10.1007/s10506-024-09420-y, <https://rdcu.be/dTIDE>
- Nicolae, Dragoș, Constantin and Yadav, Rohan, Kumar and Tufiș, Dan. Evaluation of Language Models on Romanian XQuAD and RoITD datasets. In *INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL*. vol. 18, no. 1, 2023
- Dan Cristea, Gabriela Haja , Elena Isabelle Tamba , Dan Tufiș - Digitalisation of cultures. Preparations for the adoption of a strategy for the Romanian language. In *Proceedings of the 18th International Conference on Linguistic Resources and Tools for Natural Language Processing*, pp. 23-40, 11-14 December, 2023, ISSN: 3045-2074 ISSN-L: 3045-2074
- Vasile Păiș , Verginica Barbu Mititelu , Elena Irimia , Radu Ion , Valentin Badea , Dan Tufiș , Cosmin Sterea-Grossu - Towards Anonymising the Romanian Jurisprudence. In *Proceedings of the 18th International Conference on Linguistic Resources and Tools for Natural Language Processing*, pp. 85-100, 11-14 December, 2023, ISSN: 3045-2074 ISSN-L: 3045-2074
- Dragoș Constantin Nicolae, Rohan Kumar Yadev, Dan Tufiș. A Lite Romanian BERT:ALR-BERT. In *Computers* 2022, 11, 57. MDPI Publishers, <https://doi.org/10.3390/computers11040057>
- Andrei-Marius Avram, Vasile Păiș, Dan Tufiș - Terminology annotation using a string matching algorithm In *Proceedings of the 17th International Conference on Linguistic Resources and Tools for Natural Language Processing*, pp.7-14, Chișinău, 10-12 November, 2022, ISSN: 1843-911X
- Tufiș. Dan Romanian Language Technology – a view from an academic perspective. In *INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL*. vol. 17, no. 1, 2022
- Păiș, Vasile and Ion, Radu and Avram, Andrei-Marius and Mitrofan, Maria and Tufiș, Dan. In-depth evaluation of Romanian natural language processing pipelines. In *Romanian Journal of Information Science and Technology (ROMJIST)*. vol. 24, no. 4, pp. 384--401, 2021

- Păiș, Vasile and Tufiș, Dan. Capitalization and punctuation restoration: a survey. In *Artificial Intelligence Review*. 2021
- Ungureanu, Dan and Badeanu, Madalina and Marica, Gabriela-Catalina and Dascalu, Mihai and Tufiș, Dan Ioan. Establishing a Baseline of Romanian Speech-to-Text Models. In *2021 International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*. vol. , no. , pp. 132-138, 2021
- Avram, Andrei-Marius and Păiș, Vasile and Tufiș, Dan. PyEuroVoc: A Tool for Multilingual Legal Document Classification with EuroVoc Descriptors. In *Proceedings of Recent Advances in Natural Language Processing (RANLP2021)*. pp. 92-101, 2021
- Andrei-Marius Avram, Vasile Păiș, Dan Tufiș: Towards a Romanian End-To-End Automatic Speech Recognition Based on Deepspeech2. In *PROCEEDINGS OF THE ROMANIAN ACADEMY, Series A, Vol. 21, No. 4/2020*, pp. 395–402, ISSN: 1454-9069
- Tufiș, Dan and Mitrofan, Maria and Păiș, Vasile and Ion, Radu and Coman, Andrei. Collection and Annotation of the Romanian Legal Corpus. In *Proceedings of The 12th Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, pp. 2766-2770, May 2020
- Georg Rehm, ... Tufiș, Dan, The European Language Technology Landscape in 2020: Language-Centric and Human-Centric AI for Cross-Cultural Communication in Multilingual Europe. In *Proceedings of The 12th Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, pp. 3315-3325, May 2020, ISBN: 979-10-95546-45-0.
- Tamás Váradi, ... Tufiș, Dan, The MARCELL Legislative Corpus. In *Proceedings of The 12th Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, pp. 3754-3761, May 2020, ISBN: 979-10-95546-45-0.
- Păiș, Vasile, Tufiș, Dan, Ion, Radu. MWSA Task at GlobaLex 2020: RACAI's Word Sense Alignment System using a Similarity Measurement of Dictionary Definitions. In *Proceedings of The 12th Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, pp. 69-75, May 2020, ISBN: 979-10-95546-45-0.
- Păiș, Vasile, Tufiș, Dan, Ion, Radu. A Processing Platform Relating Data and Tools for Romanian Language. In *Proceedings of The 12th Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, pp. 81-88, May 2020, ISBN: 979-10-95546-45-0.
- Radu Ion, Valentin Gabriel Badea, George Cioroiu, Verginica Barbu Mititelu, Elena Irimia, Maria Mitrofan, Dan Tufiș (2020). A Dialog Manager for Micro-Worlds. *Studies in Informatics and Control*, 29 (4), 401-410, ISSN 1220-1766, 2020
- Avram, A.M., Păiș, V., Tufiș, D. (2020). Romanian Speech Recognition Experiments from the ROBIN Project. In *Proceedings of the 15th International Conference on Linguistic Resources and Tools for Natural Language Processing*, pp. 103-114, 14-16 December, ISSN: 1843-911X, 2020
- Kupietz, M., Diewald, N., Trawiński, B., Cosma, R., Cristea, D., Tufiș, D., Váradi, T., Wöllstein, A. (2020.: Recent developments in the European Reference Corpus EuReCo, In Sylviane Granger & Marie-Aude Lefer (eds) *Translating and Comparing Languages: Corpus-*

- based Insights*. Corpora and Language in Use Proceedings 6, Louvain-la-Neuve: Presses universitaires de Louvain pp.257—276, July 2020
- Tufiș, D., Mitrofan, M. Păiș, V., Ion, R. Coman, A. (2020). Collection and Annotation of the Romanian Legal Corpus. In *Proceedings of The 12th Language Resources and Evaluation Conference*. European Language Resources Association, Marseille, France, pp. 2766-2770, May 2020
 - Tufiș, D., Barbu Mititelu, V., Irimia, E., Păiș, V., Ion, R., Diewald, N., Mitrofan, M., Onofrei, M. (2019). Little Strokes Fell Great Oaks. Creating Corola, The Reference Corpus of Contemporary Romanian. *Revue roumaine de linguistique*, In *Revue roumaine de linguistique*, No./Issue 3, 2019
 - Coman, A., Mitrofan, M., Tufiș, D. (2019). Automatic Identification and Classification of Legal Terms in Romanian Law Texts. In Proceedings of the 14th Edition of the International Conference on Linguistic Resources and Tools for Natural Language Processing, Iași, Noiembrie, pp.39-49, ISSN 1843-911X, 2019
 - Păiș, V., Tufiș, D., Ion, R. (2019). Integration of Romanian NLP tools into the RELATE platform. In Proceedings of the 14th Edition of the International Conference on Linguistic Resources and Tools for Natural Language Processing, Iași, Noiembrie, pp.181-192, ISSN 1843-911X, 2019
 - Skandina, I., Gaizauskas, R., Babych, B., Ljubescic, N., Tufiș, D., Vasiljevs, A. (2019) (eds). Using Comparable Corpora for Under-Resourced Areas of Machine Translation. In *series Theory and Applications of Natural Language Processing*, Springer, 978-3-319-99003-3, 2019
 - Aker, A., Ceaușu, A., Feng, Y., Gaizauskas, R., Hunsicker, S., Ion, R., Irimia, E., Ștefănescu, D., Tufiș, D. (2019). Mapping and aligning units from comparable corpora. În Inguna Skandina, Robert Gaizauskas, Bogdan Babych, Nicola Ljubescic, Dan Tufiș, Andrejs Vasiljevs (eds). Using Comparable Corpora for Under-Resourced Areas of Machine Translation. In *series Theory and Applications of Natural Language Processing*, Springer, 978-3-319-99003-3, 2019
 - Lestari Paramita, M., Aker, A., Clough, P., Gaizauskas, R., Glaros, N., Mastropavlos, N., Yannoutsou, O., Ion, R., Ștefănescu, D., Ceaușu, A., Tufiș, D. (2019). Collecting Comparable Corpora. În Inguna Skandina, Robert Gaizauskas, Bogdan Babych, Nicola Ljubescic, Dan Tufiș, Andrejs Vasiljevs (eds). Using Comparable Corpora for Under-Resourced Areas of Machine Translation. In *series Theory and Applications of Natural Language Processing*, Springer, 978-3-319-99003-3, pp. 55-88, 2019
 - Dan Tufiș, Verginica Barbu Mititelu, Elena Irimia, Maria Mitrofan, Radu Ion, George Cioroiu. (2019). Making Pepper Understand and Respond in Romanian. Proceedings of the 22nd International Conference on Control Systems and Computer Science, București, 2019