



ACADEMIA ROMÂNĂ

SECȚIA DE ȘTIINȚA ȘI TEHNOLOGIA INFORMAȚIEI

Institutul de Cercetări pentru Inteligență Artificială „Mihai Drăgănescu”

Nr. ICIA: 348/28 noiembrie 2018

RAPORT DE ACTIVITATE
Acad. Ioan Dan Tufiș
PE ANUL 2018

CASA ACADEMIEI -- Calea 13 Septembrie 13, București 5, ROMÂNIA, 050711

Tel: + (4021) 3188103, Fax: +(4021) 3188142



ACADEMIA ROMÂNĂ

SECȚIA DE ȘTIINȚA ȘI TEHNOLOGIA INFORMAȚIEI

Institutul de Cercetări pentru Inteligență Artificială „Mihai Drăgănescu”

Rezultatele cercetării desfășurate în anul 2018

Coordonator al PROGRAMULUI PRIORITAR DE CERCETARE: *Corpus computațional de referință pentru limba română contemporană* cu două faze:

Faza 1: *Eliminare erori, completare meta-informații lipsă, adăugarea de noi texte*
Faza a 2-a: *Migrarea COROLA pe platforma KoRaP*

Activitate formare de tineri cercetători: doctoranzi

Din cei 6 doctoranzi ai ICIA, 4 sunt în coordonarea Acad. Ioan Dan Tufiș.

Articole publicate într-o revistă cotate de *Web of Science* (Thomson Reuters)

1. Dan Tufiș, Dan Cristea A Bird's-eye View of Language Processing Projects at Romanian Academy, Miyazaki, Japan, LREC 2018, pp. 2445-2451, ISBN 979-10-95546-00-9
2. Maria Mitrofan, Dan Tufiș: BioRo: The Biomedical Corpus for the Romanian Language, Miyazaki, Japan, LREC 2018, pp. 1192-1196, ISBN 979-10-95546-00-9
3. Verginica Barbu Mititelu, Dan Tufiș, Elena Irimia: The Reference Corpus of the Contemporary Romanian Language (CoRoLa), Miyazaki, Japan, LREC 2018, pp. 1178-1185, ISBN 979-10-95546-00-9

Articole publicate in reviste ISI în România

1. Vasile Păiș, Dan Tufiș,. Computing distributed representations of words using the CoRoLa corpus. In Proceedings of the Romanian Academy, series A, pp. 403-410, 2018

Capitole in Cărți apărute într-o editură consacrată din străinătate

1. Barbu Mititelu, Verginica and Irimia, Elena and Tufiș, Dan. Linking the Romanian Wordnet to the Open Multilingual Wordnet. Special Issue on Linking, Integrating and Extending Wordnets. In *CSLI Publications, Stanford*. 2018 (in print)

CASA ACADEMIEI -- Calea 13 Septembrie 13, București 5, ROMÂNIA, 050711

Tel: + (4021) 3188103, Fax: +(4021) 3188142



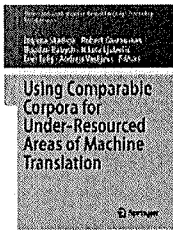
ACADEMIA ROMÂNĂ

SECȚIA DE ȘTIINȚA ȘI TEHNOLOGIA INFORMAȚIEI

Institutul de Cercetări pentru Inteligență Artificială „Mihai Drăgănescu”

- Ahmet Aker, Alexandru Ceașu, Yang Feng, Robert Gaizauskas, Sabine Hunsicker, Radu Ion, Elena Irimia, Dan Ștefănescu and Dan Tufiș. Mapping and aligning units from comparable corpora. În Skandina, Inguna and Gaizauskas, Robert and Babych, Bogdan and Ljubesic, Nicola and Tufiș, Dan and Vasiljevs, Andrejs. Using Comparable Corpora for Under-Resourced Areas of Machine Translation. In *series Theory and Applications of Natural Language Processing, Springer, 2018-2019, 978-3-319-99003-3*

Cărți publicate în străinătate



Skandina, Inguna and Gaizauskas, Robert and Babych, Bogdan and Ljubesic, Nicola and Tufiș, Dan and Vasiljevs, Andrejs. *Using Comparable Corpora for Under-Resourced Areas of Machine Translation*. In series Theory and Applications of Natural Language Processing, Springer, 2018-2019, 978-3-319-99003-3 9

Cărți publicate în țară

- Daniela Gîfu, Diana Trandabăț, Vasile Păiș, Dan Cristea, Dan Tufiș (eds) The Proceedings of the 13th Edition of the **International Conference on Linguistic Resources and Tools for Processing Romanian Language.**, Iași, Noiembrie 2018 (in print) Volumul urmează să fie indexat de Springer Verlag

Capitole în cărți publicate în țară

- Vasile Păiș, Dan Tufiș. **More Romanian Word Embeddings from the ReTeRom Project**

Citări ale lucrărilor anterioare în 2018

- În 2018 am 105 de citări (dintr-un total de 2997) – cf. Google Scholar

CASA ACADEMIEI -- Calea 13 Septembrie 13, București 5, ROMÂNIA, 050711

Tel: + (4021) 3188103, Fax: +(4021) 3188142



ACADEMIA ROMÂNĂ

SECȚIA DE ȘTIINȚA ȘI TEHNOLOGIA INFORMAȚIEI

Institutul de Cercetări pentru Inteligență Artificială „Mihai Drăgănescu”

Rapoarte de interes public

- CoRoLa-based Word Embeddings (rezultatele prezentate în acest raport și în articolul publicat de Proc. of RomAc sunt utilizate deja cel puțin de specialiști de la IDS-Mannheim Germania, de la Universitatea Babeș-Bolyai și citate în articolele scrise de ei)

Cea/cele mai important/e realizare/i:

- Proiectul CoRoLA în jurul căruia gravitează mai multe proiecte naționale (ReTeRom, ROBIN) și europene (MARCELL, ELRC, ELG)
- Volumul co-editat și apărut la Springer *Using Comparable Corpora for Under-Resourced Areas of Machine Translation*. In series Theory and Applications of Natural Language Processing, Springer, 2018-2019, 978-3-319-99003-3 9

Conferințe organizate

1. 6th Int. Workshop on Semantic and Collaborative Technologies for the Web, in conjunction with the 13th Int. Conf. on eLearning and Software for Education (eLSE 2018)
2. A 13^a ediție a Conferinței internaționale Resurse Lingvistice și Instrumente de prelucrare a limbii române (ediție dedicată Centenarului).
3. Workshopul Corpus Harmonization and Language Comparison (în cadrul proiectului DRUKOLA), Facultatea de Limbi Străine din Universitatea București

Participări la conferințe

1. Dan Tufiș, Alexandru Ceaușu **Platforma eTranslation a MIE @ lucru**, la European Commission Workshop ELRC
 2. Dan Tufiș **ELRC în Romania**, la European Commission Workshop ELRC
 3. Vasile Păiș, Dan Tufiș, "CoRoLa-based Word Embeddings", la Workshopul Corpus Harmonization and Language Comparison (în cadrul proiectului DRUKOLA), Facultatea de Limbi Străine din Universitatea București
-

CASA ACADEMIEI -- Calea 13 Septembrie 13, București 5, ROMÂNIA, 050711

Tel: + (4021) 3188103, Fax: +(4021) 3188142



ACADEMIA ROMÂNĂ

SECȚIA DE ȘTIINȚA ȘI TEHNOLOGIA INFORMAȚIEI

Institutul de Cercetări pentru Inteligență Artificială „Mihai Drăgănescu”

4. Tufiș, D., Cristea, D., CoRoLA (<http://www1.ids-mannheim.de/kl/projekte/drukola/workshop-corpus-harmonization-and-language-comparison.html>)

Cooperări științifice naționale și internaționale, inclusiv în cadrul proiectelor:

În ICIA în anul 2018 s-au derulat sub coordonarea mea următoarele proiecte:

Proiecte naționale:

- **ReTeRom** - Resurse și tehnologii pentru dezvoltarea interfețelorom-mașina în limba română PN-III-P1-1.2-PCCDI-2018-0818, nr. 73PCCDI/2018, 2018-2020, director de proiect complex: Tufis, D, total: 1253617 lei din care în 2018: 359779 lei
- **ROBIN** – Roboții și Societatea: Sisteme Cognitive pentru Roboți Personali și Vehicule Autonome PN-III-P1-1.2-PCCDI-2018-0818, nr. 72PCCDI/2018, 2018-2020, responsabil ICIA: Tufis, D. Total: 950500 lei din care în 2018: 270400 lei
- **CoRoLa** – Proiect Prioritar al Academiei Române 2017-2019, coordinator: Tufis, D.

Proiecte internaționale:

- **DRuKoLA** - Sprachvergleich korpustechnologisch. Deutsch – Rumänisch (funded by the Alexander von Humboldt-Foundation), <http://www1.ids-mannheim.de/direktion/kl/projekte/drukola.html>, 2016 - 2018; total: 25000 Euro din care în 2018: 8700 EURO
- **ELRC** – European Language Resource Coordination (CEF SMART 2014/1074)– Tufis, D. este National Technological Anchor Point – coordonator si membru în consiliul de coordonare al proiectului , 2016-2020: Tufis, D., total: 28200 EURO din care în 2018: 17400 EURO
- **MARCELL** - "Multilingual Resources for CEF.AT in the legal domain" (INEA/CE F/ICT/A2017/1565710), 2018-2020, coordonator ICIA: Tufis D. total: 322922,8 EURO din care în 2018: 27156,73 EURO

Proiecte internaționale de anvergura aflate în faza de evaluare sau de pregătire în care este implicat ICIA:

CASA ACADEMIEI -- Calea 13 Septembrie 13, București 5, ROMÂNIA, 050711

Tel: + (4021) 3188103, Fax: +(4021) 3188142



ACADEMIA ROMÂNĂ

SECȚIA DE ȘTIINȚA ȘI TEHNOLOGIA INFORMAȚIEI

Institutul de Cercetări pentru Inteligență Artificială „Mihai Drăgănescu”

- **CEF Automated Translation Toolkit for the EU Council Presidency**
- **EUREKO** – European Reference Corpus (vezi lucrarea Recent developments in the European Reference Corpus (EuReCo) prezentată și la conferința Using Corpora in Contrastive and Translation Studies (5th edition) <https://uclouvain.be/en/research-institutes/ilc/cecl/uccts2018presentations.html>)
- **NexusLinguarum** – European Network for Research and Development in Linguistic Data Science
- **ELG** -European Language Grid
- **HLP** -Human Language Project (o inițiativa pentru un proiect de tip ”flagship”)

Număr de citări:

- În 2018 am 105 de citări (dintr-un total de 2997) – cf. Google Scholar

Membru în conducerea unei organizații internaționale de specialitate

- Board of the ”European Language Resource Coordination (ELRC)” proiect: 2015-2020

Apartenență la comitete de program al unor conferințe internaționale:

- LREC 2018 - 11th edition of the Language Resources and Evaluation Conference, 7-12 May 2018, Miyazaki (Japan)

Membru în colectivul de redacție al unei reviste naționale/internaționale (cotată de *Web of Science*, Thomson Reuters sau indexată într-o BDI) sau în colectivul editorial al unor edituri internaționale consacrate

- International Journal on Language Resources and Evaluation, Springer Verlag, USA - ISI
 - Proceedings of the Romanian Academy, Series A (din 1999)- topical editor ISI
-

CASA ACADEMIEI -- Calea 13 Septembrie 13, București 5, ROMÂNIA, 050711

Tel: + (4021) 3188103, Fax: +(4021) 3188142



ACADEMIA ROMÂNĂ

SECȚIA DE ȘTIINȚA ȘI TEHNOLOGIA INFORMAȚIEI

Institutul de Cercetări pentru Inteligență Artificială „Mihai Drăgănescu”

- Romanian Journal on Science and Technology of Information, Academia Română ISI
- Studies in Informatics and Control, România ISI
- International Journal of Computers, Communication & Control, Universitatea „Agora” din Oradea ISI
- Romanian Journal of Human – Computer Interaction, ACM SIGCHI Romania -BDI
- NOESIS – Travaux du comite roumain d’histoire et de philosophie de sciences, Academia Română (director) - BDI
- The Computer Science Journal of Moldova, Republica Moldova –BDI
- Translation and Multilingual Natural Language Processing, Language Science Press

Alte realizări excelente

- Evaluator Comisia Europeana FET-OPEN-RIA (2016-2018)
- Ales (2017-2020) în Colegiul de Evaluatori Experti ai European Science Foundation
- Membru (președinte) în Comisia de selecție a candidaților pentru comisia de informatică a CNATCDU
- Membru de onoare (*fellow, senior*) ai unei societăți științifice naționale/internaționale: Ioan Dan Tufiș (ACM)

Acad. Ioan Dan Tufiș

CASA ACADEMIEI -- Calea 13 Septembrie 13, București 5, ROMÂNIA, 050711

Tel: + (4021) 3188103, Fax: +(4021) 3188142