

## RAPORT DE ACTIVITATE PE ANUL 2010

Prof. Dan Tufiş

Am condus, în calitate de director, activitatea științifică și managerială a Institutului de Cercetări pentru Inteligență Artificială al Academiei Române.

Sunt membru în Biroul Secției de Știința și Tehnologia Informației din Academia Română.

Am condus activitatea de doctorat din institut, doi dintre doctoranzii mei (Dan Ștefănescu și respectiv Nadia Luiza Dincă) susținându-i cu succes teza de doctorat (Dan Ștefănescu -Magna Cum Laudae). Alți doi doctoranzi la ICIA și-au susținut (în luna decembrie 2010) public tezele elaborate sub conducerea Acad. Fl. Gh. Filip (Ana Suduc, Mihai Bizoi).

Am făcut parte din 11 comisii de doctorat la Universitatea A.I. Cuza și la Academia Română.

În luna decembrie mi s-a acordat titlul de „**Profesor de Onoare**” al Universității „A.I. Cuza” Iași. De menționat că acest titlu a fost până acum oferit de facultatea de Informatică doar unui număr de 4 personalități, eu fiind singurul laureat din România.

Tot în decembrie 2010 am primit diploma de „**Professional Membership**” a ACM (Association for Computing Machinery).

Am făcut 18 recenzii pentru o parte din revistele internaționale și naționale din al căror Comitet editorial fac parte:

- International Journal on Language Resources and Evaluation, Springer Verlag, USA
- Journal on Computers and Humanities, Springer Verlag, USA
- Book Series “Text and Language Technology”, Springer Verlag, USA
- International Journal of Corpus Linguistics, John Benjamin Publishers, UK
- The Computer Science Journal of Moldova, Republica Moldova
- Proceedings of the Romanian Academy, Series A
- Romanian Journal on Science and Technology of Information, Academia Română
- International Journal on Information and Control, ICI, România
- International Journal of Computers, Communication & Control, Universitatea „Agora” din Oradea
- Romanian Journal of Human – Computer Interaction, ACM SIGCHI Romania

Am coordonat participarea institutului, în calitate de responsabil sau director de proiect, în următoarele proiecte:

1. FlaReNet (2008-2010) Proiect e-ContentPlus cu 28 de parteneri
2. 2008-2018 Common Language Resource Infrastructure (CLARIN), proiect FP7-Infrastructure cu 30 de parteneri.
3. ACCURAT (2010-2012) Proiect IST-FP7 de traducere automată cu 7 parteneri, în care teoria reificării, elaborată de mine în anul 2007, are un loc central.

4. MULTILINGUAL WEB (2010-2012) Proiect european IST-PSP. cu 10 parteneri, de promovare a standardelor multilinguale și de definire a unora noi, pentru asigurarea interoperabilității instrumentelor de prelucrare a documentelor de pe web.
5. Meta-Net (Multilingual Europe: a Technology Alliance) (2009-2011) Proiect FP7-NoE, 249119 - Network of Excellence, coordonat de DFKI Germania și Meta-Net4You, Proiect IST-PSP satelit al META-NET care va începe în 2011 (proiect cu 8 parteneri, pentru construcția de resurse și instrumente lingvistice în conformitate cu specificațiile META-NET)
6. SIR-RESDEC, proiect PN2 (2007-2010) încheiat la 15 septembrie 2010
7. eDTLR, proiect PN2 (2007-2010) încheiat la 15 decembrie 2010
8. STAR (2008-2011), proiect CNCISIS

Proiectul FP7 de infrastructură de cercetare CLARIN (sunt vicepreședintele Comitetului Științific al proiectului) a cărui etapă pregătitoare s-a încheiat în decembrie 2010 a fost extins (fără finanțare până în iunie 2011). Pe lângă numeroasele specificații de construcție, CLARIN a creat sub coordonarea mea un prototip de demonstrație (European CLARIN Demonstrator). Proiectul se află într-o fază avansată de creare a instituției ERIC-CLARIN (România a semnat deja două angajamente ERIC pentru proiectele ELI și FAIR). Ca reprezentant CLARIN pentru România am făcut o serie de prezentări la Agenția Națională de Cercetare Științifică și am elaborat o serie de documente pregătitoare pentru aderarea României la ERIC-CLARIN (fazele de construcție și exploatare a infrastructurii de cercetare pan-europeană CLARIN).

Am fost membru în comitetul de organizare (CO), comitetul de program (CP) și/sau comitetul de recenzie (CR) făcând peste 50 de recenzii de lucrări la conferințele următoare:

1. LREC 2010, 18-23 mai 2010, Malta, membru în CP, CR, președinte de secțiune
2. Workshop on Language Technology as Web Services, Malta, 2010, membru în CO, CP, CR
3. Workshop on Language Resource and Language Technology Standards – state of the art, emerging needs, and future developments, Malta, 2010, membru în CP, CR
4. A 18a Conferință internațională „Intelligent Information Systems” (IIS 2010), Siedlce, Polonia, 8-10 iunie 2010 membru în CP, CR
5. AIMSA 2010, Varna, Bulgaria, 8-10th September, 2010, membru în CP, CR
6. Workshop on Exploitation of multilingual resources and tools for Central and (South) Eastern European Languages, Malta, 2010, membru în CP, CR
7. Workshop on Semantic Relations. Theory and Applications, Malta, 2010, membru în CP, CR
8. 1<sup>st</sup> Workshop on the Analysis of Subjectivity and Sentiment focused on a Unified Practice (WASSUP), European Conference on Artificial Intelligence, Lisbon, Portugal 2010, membru în CP, CR
9. Digital Humanities 2010, London, 2010, membru în CR
10. 2nd Workshop on „Cognitive Aspects of the Lexicon”, The 23rd International Conference on Computational Linguistics (COLING 2010), Beijing, China on August 23-27th, 2010, membru în CP, CR
11. SYNASC 2010 - 12th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing, Timisoara, Romania, September 23-26, 2010, membru în CP, CR
12. COLING 2010 - 23rd International Conference on Computational Linguistics, Beijing, China, August 23-27, 2010, membru în CR
13. SOFA 2010 - 4th IEEE International Workshop on Soft Computing Applications, Arad, Romania, July 15-17, 2010, membru în CR

14. NLPCS 2010 - 7th International Workshop on *Natural Language Processing and Cognitive Science* at the 12th International Conference on Enterprise Information Systems - ICEIS 2010, Madeira, Portugal, June 8-7, 2010, membru în CP, CR
15. EAMT 2010 - 14th Annual Conference of the European Association for Machine Translation, Saint-Raphaël, France, May 27-28, 2010, membru în CR
16. PROMISE2010 (Processing Romanian in Multilingual, Interoperational and Scalable Environments), Universitatea „A.I. Cuza” Iași, 29-31 martie, 2010, membru în CO, CP, CR
17. ConsILR-2010 - Conferința *Resurse Lingvistice și Instrumente pentru Prelucrarea Limbii Române*, București, Romania, 6-7 mai, 2010, membru în CO, CP, CR
18. CCCT 2010 - 8th International Conference on Computing, Communications and Control Technologies, Florida, USA, April 6-9, 2010, membru în CP, CR
19. MLQA 2010 - Workshop on Question Answering in Multiple Languages, Padua, Italy, September 22, 2010, membru în CP, CR
20. Multilingual Web 2010 - W3C Workshop: The Multilingual Web - Where Are We?, Madrid, Spain, October 27-28, 2010, membru în CP, CR
21. EMNLP 2010 - 15th Conference on Empirical Methods in Natural Language Processing, MIT, Massachusetts, USA, October 9-11, 2010, membru în CP, CR

Am fost co-organizator la trei conferințe de referință ale domeniului:

22. Workshop on Language Technology as Web Services, Malta, 28-30 mai 2010
23. Conferința Internațională– Resurse și instrumente de prelucrare automată a limbii române, Muzeul literaturii române, București, 5-6 mai 2010
24. Workshopul Internațional PROMISE (Processing Romanian in Multilingual, Interoperational and Scalable Environments), Universitatea „A.I. Cuza” Iași, 29-31 martie, 2010
25. Workshopul Multilingual Web, ICIA, Academia Română, București, 30 mai-1 iunie, 2010

Am publicat o serie de lucrări (volume editate, capitole de carte, articole la conferințe și articole în reviste de specialitate) după cum urmează:

### **I.1. Volume de autor sau editate**

26. Dan Tufiș, Corina Forăscu (eds.): *Multilinguality and Interoperability in Language Processing with Emphasis on Romanian*. Editura Academiei, București, 2010, ISBN 978-973-27-1972-5, 478 pages
27. Adrian Iftene, Horia N. Teodorescu, Dan Cristea, Dan Tufiș (eds.): *Resurse lingvistice și instrumente pentru prelucrarea limbii române*. Editura Universității “A. I. Cuza”, ISSN 1843-911X, 238 pages, 2010

### **I.2. Capitole de carte**

28. Inguna Skadina, Ahmet Aker, Voula Giouli, Dan Tufiș, Robert Gaizauskas, Madara Mieriņa and Nikos Mastropavlos. “Collection of comparable corpora for under-resourced languages”. In *Frontiers in Artificial Intelligence and Applications*. IOS Press, Vol. 219, ISBN 978-1-60750-640-9, pages 161-168, 2010 (ISI Proceedings)

29. Dan Tufiş, Radu Ion, Alexandru Ceauşu, Dan Ştefănescu: Reifying the Alignments. In Tufiş, D., Corina Forăscu (eds.): *Multilinguality and Interoperability in Language Processing with Emphasis on Romanian*. Editura Academiei, pp. 69-86, 2010.
30. Tufiş, D.: Part of Speech Tagging. In Mitkov, R (ed.) *Handbook of Computational Linguistics*, Oxford Press, 2010 (in press)
31. Ştefănescu, D., Ion, R., Ceauşu, A., Tufiş, D.: Sistem întrebare-răspuns antrenabil pentru limba română. În Adrian Iftene, Horia N. Teodorescu, Dan Cristea, Dan Tufiş (eds.): *Resurse lingvistice și instrumente pentru prelucrarea limbii române*. Editura Universităţii “A. I. Cuza”, ISSN 1843-911X, pp. 145-156, 2010
32. Boroş, T., Tufiş, D., Ceauşu, A.: Construcţia automată de corpusuri multilinguale. În Adrian Iftene, Horia N. Teodorescu, Dan Cristea, Dan Tufiş (eds.): *Resurse lingvistice și instrumente pentru prelucrarea limbii române*. Editura Universităţii “A. I. Cuza”, ISSN 1843-911X, pp. 95-104, 2010
33. Ion, R., Ceauşu, A., Ştefănescu, D., Tufiş, D.: Servicii web interoperabile și multilinguale. În Adrian Iftene, Horia N. Teodorescu, Dan Cristea, Dan Tufiş (eds.): *Resurse lingvistice și instrumente pentru prelucrarea limbii române*. Editura Universităţii “A. I. Cuza”, ISSN 1843-911X, pp. 167-176, 2010
34. Radu Ion, Dan Ştefănescu, Alexandru Ceauşu, Dan Tufiş, Elena Irimia, Verginica Barbu-Mititelu: A Trainable Multi-factored QA System. In Carol Peters, Giorgio Di Nunzio, Mikko Kurimo, Thomas Mandl, Djamel Mostefa, Anselmo Peñas, Giovanna Roda (eds.) *Multilingual Information Access Evaluation, Vol. I Text Retrieval Experiments*, pp. 257-264, Lecture Notes in Computer Science, Volume 6241/2010, Springer-Verlag. ISBN: 978-3-540-85759-4. (ISI Proceedings)

### **I.3. Contribuții în volume de lucrări (conference proceedings)**

35. R. Ion, D. Tufiş, T. Boroş, A. Ceauşu, D. Ştefănescu: On-line Compilation of Comparable Corpora and their Evaluation. In *Proceedings of the 7<sup>th</sup> Conference on Formal Approaches to South Slavic and Balkan Languages (FASSBL 2010)*, Dubrovnic, October, 2010, pp 29-34.
36. Inguna Skadiņa, I. Vasiļjevs, A., Skadiņš, R., Gaizauskas, R., Tufiş, D. Gornostay, T.: Analysis and Evaluation of Comparable Corpora for Under Resourced Areas of Machine Translation. In *Proceedings of the 3rd Workshop on Building and Using Comparable Corpora, LREC 2010, Malta, 22 May 2010*, pages 6–14
37. Tufiş D., Ştefănescu D.: A Differential Semantics Approach to the Annotation of the Synsets in WordNet. In *Proceedings of LREC 2010, Malta, May 2010*, ELRA, ISBN 2-9517408-6-7
38. P. Wittenburg, E. Hajicova, G. Budin, K. Koskenniemi, M. Wynne, T. Váradi, L. Lemnitzer, N. Calzolari, L. Borin, M. Piasecki, N. Bel, S. Piperidis, D. Tufiş, R.v. Veenendaal, B. Maegaard, J.M. Pierrel, I. Skadina: Resource and Service Centre's as the Backbone for a Sustainable Service Infrastructure. In *Proceedings of LREC 2010, Malta, May 2010*, ISBN 2-9517408-6-7
39. Radu Ion, Alexandru Ceauşu, Dan Ştefănescu, Dan Tufiş, Elena Irimia and Verginica Barbu Mititelu. Monolingual and Multilingual Question Answering on European Legislation. In Martin Braschler, Donna Harman, and Emanuele Pianta (eds.) *CLEF 2010 LABs and Workshops (Notebook Papers)*, Padua, Italy, September 2010. ISBN: 978-88-904810-0-0, ISSN: 2038-4963.

#### I.4. Lucrări publicate sau trimise spre publicare în periodice

40. D. Tufiș, D. Ștefănescu: Annotation of Wordnet towards Multifactorial Textual Analysis, 20 pagini, (submitted).
41. Tufiș, D.; Ion, R.; Ceaușu, A.; Ștefănescu, D.; Irimia, E.; Mititelu, V. (2010), « *Servicii Web lingvistice* », în *Revista CALITATEA acces la succes*, nr. 10, 2010, pg. 60 – 63, ISSN : 1582-2559
42. Tufiș, D. (2010), “Inteligența Artificială elimină “zgomotul informațional” din mediul online”, în *Market Watch*, nr. 129, oct-nov, 2010, pg. 42 – 43.

#### I.5 Conferințe invitate

43. Dan Tufiș: A differential semantics method for annotating the WordNet synsets. Mac Plank Institute, Neijmegen, 18 ianuarie, 2010
44. Dan Tufiș: *Multi-factor Lexical Mark-up of Connotation in a WordNet Lexical Ontology*, Categorizing Human Experience: Classification in Languages and Knowledge Systems, Ecole Normale Supérieure (ENS), Paris, France May 14-15, 2010

În perioada octombrie-decembrie am participat substanțial la redactarea unei propuneri de proiect european în cadrul FP7 (Spark-MT) depus pe data de 18 ianuarie 2011 la Comisia Europeană. Proiectul promovează modele, tehnici și soluții originale dezvoltate în grupul de cercetare pe care îl conduc, ICIA fiind un partener strategic al proiectului.

Publish or Perish ([http://www.harzing.com/pop.htm?source=pop\\_3.1.4004#download](http://www.harzing.com/pop.htm?source=pop_3.1.4004#download)), cel mai popular instrument gratuit de analiză bibliometrică, arăta la începutul anului 2011 un număr de **1350 citări** ale lucrărilor mele indexate de Google Scholar (h-index17).