

## CANDIDAT PENTRU ALEGEREA CA MEMBRU TITULAR

---

**Numele candidatului:** Tufiș Ioan Dan

**Locul de muncă și funcția:** Institutul de Cercetări pentru Inteligență Artificială, Academia Română, Director

**Recomandat de:** Academician Gheorghe Tecuci

**Data și locul nașterii:** 5 februarie 1954, București

**Adresa prezentă:** Str. Lipova, nr. 41, sector 1, 010888, București

---

### *Justificarea propunerii*

Prof. Tufiș, o personalitate științifică recunoscută și stimată pe plan mondial, a definit și implementat concepte și modele originale de tratare a unor probleme dificile în prelucrarea automată a limbajului naturală (rețele semantice sensibile la navigare, învățare automată a morfologiei paradigmatică a unei limbi naturale, dezambiguizarea morfo-sintactică ierarhizată, alinierea reificată a textelor bilingve, construcția ontologiilor lexicale, localizarea pe web a răspunsurilor relevante la întrebări arbitrare, semantica multifactorială etc.). Cercetările sale s-au concretizat prin realizarea unor sisteme inteligente de cel mai înalt nivel, câștigătoare în competiții internaționale cu cele mai bune sisteme din lume dezvoltate la Stanford Univ., Carnegie Mellon Univ., MIT Research, Univ. of California, Xerox Research Europe, Microsoft Research etc. Peste 230 dintre lucrările sale sunt indexate în principalele baze de date internaționale (ACL Anthology, ISI, Google Scholar, ACM, Scopus, CiteSeer, DBLP, Computer Science Bibliographies, MSN, Scirus, ScienceDirect etc.). Programul scientometric "Publish or Perish" indexează 230 de lucrări dintre care 159 sunt citate în total de 1500 ori, cu factorul h-index 19 (<http://www.harzing.com/pop.htm#download>). A primit Premiul Academiei Române în 1989 și 1994. Prof. Tufiș a creat o școală românească de prestigiu internațional în domeniul prelucrării automate a limbajului natural. A înființat Comisia de Informatizare pentru Limba Română al cărui președinte este din anul 2001, for național cu cercetători din țară și din străinătate. Este membru în comitetele de redacție a 13 publicații internaționale (7 din străinătate și 6 din țară). A susținut peste 50 de conferințe invitate în SUA, Europa și Asia. În decembrie 2010 a primit titlul de "Profesor de Onoare" al Universității "A. I. Cuza" din Iași (acest titlu a fost până acum acordat de Facultatea de Informatică doar unui număr de 4 personalități, Dan Tufiș fiind singurul laureat din România).

În perioada 1990-2011, prin rezultatele sale, Dan Tufiș a câștigat 36 de proiecte internaționale de cercetare și 10 naționale, ce au susținut cercetarea românească cu peste 4.000.000 euro.

### *Realizări cu impact internațional (după primirea în Academia Română - listă parțială)*

**Modelul ierarhic al dezambiguizării automate morfo-sintactice** („tiered tagging” publicat în 1999): *una dintre cele mai apreciate abordări din lume în prelucrările morfo-lexicale de tip statistic.*

**TREQ-AL și COWAL: Cele mai bune sisteme de aliniere lexicală pentru traducere automată.** La Conferința Asociației Nord-Americane de Lingvistică Computațională, Edmonton, Canada, 2003, *TREQ-AL câștigă prima competiție de aliniere lexicală* devansând 12 sisteme ale unor universități din SUA și Canada. Între anii 2003-2005 elaborează *teoria alinierii prin reificare*, folosită în sistemul COWAL care, în 2005, la Ann Arbor, Michigan, SUA, *câștigă și a doua competiție de aliniere lexicală* la concurență cu 37 de sisteme de la universități și companii de prestigiu din lume.

**Câștigător în competițiile europene CLEF 2006, CLEF 2007 și CLEF 2009 cu sisteme de întrebare-răspuns în spațiul web.**

A definit și implementat o teorie de semantică diferențială originală pentru cuantificarea denotației lexicale (o amplă lucrare despre acest model va apărea în 2012 în revista "Decision Support Systems", revistă ISI cu un factor de impact de 2,135 publicată de Elsevier).

Din 2001 *coordonează Ontologia Lexicală RO-WordNet, lucrare monumentală de lexicografie computațională* aliniată la nivel conceptual cu ontologia lexicală WordNet începută în SUA, la Universitatea Princeton, cu peste 30 de ani în urmă.

15 noiembrie, 2011

Acad. Florin Gh. Filip  
Președinte al Secției "Știința și Tehnologia Informației"

---

**Decizia Secției:** Propunere aprobată în unanimitate, prin vot secret, de toți membrii titulari ai Secției, la data de 14 februarie 2008

**5 voturi exprimate; 5 voturi pentru; 0 voturi contra; 0 voturi abținere**

---

## 12 Lucrări științifice cu impact internațional (după primirea în Academia Română)<sup>1</sup>

1. R. Steinberger, B. Pouliquen, A. Widiger, C. Ignat, T. Erjavec, **D. Tufiș**, Daniel Varga. The JRC-Acquis: A multilingual aligned parallel corpus with 20+ languages. In Proc. of the 5<sup>th</sup> LREC Conf., Genoa, Italy, 22-28 May, 2006, pp.2142-2147, ISBN 2-9517408-2-4 (**178 citări**).  
*Lucrarea descrie procesul de creare și aliniere a corpusului paralel JRC-Acquis, corpusul fundamental de antrenare a sistemelor statistice de traducere automată.*
2. N. Ide, T. Erjavec, **D. Tufiș**: Sense Discrimination with Parallel Corpora. In Proc. of the SIGLEX Workshop on Word Sense Disambiguation: Recent Successes and Future Directions. Proc. of ACL2002, Philadelphia 2002, pp. 56-60 (**81 citări**).  
*Lucrarea descrie o metodă foarte eficientă și algoritmi asociați introduși de D.Tufiș pentru alinierea lexicală a textelor paralele.*
3. L. Dimitrova, T. Erjavec, N. Ide, H.J. Kaalep, V. Petkevic, **D. Tufiș**: Multext-East: Parallel and Comparable Corpora and Lexicons for Six Central and Eastern European Languages, In Proc. of COLING, Montreal 1998 (**79 citări**).  
*Lucrarea descrie metode de reprezentare și prelucrare a corpusurilor monolingve care s-au impus ca standarde de facto internaționale ale domeniului..*
4. **D. Tufiș**, O. Mason: Tagging Romanian Texts: A Case Study for QTAG, a Language Independent Probabilistic Tagger, Proc of 1<sup>st</sup> Conf. on Language Resources and Evaluation, Granada, May, 1998, pp. 589-596 (**72 citări**)  
*Prezintă unul dintre primele dezambiguizatoare morfo-lexicale, bazate pe un model pur statistic.*
5. **D. Tufiș**, D. Cristea, S. Stamou. BalkaNet: Aims, Methods, Results and Perspectives. A General Overview. In Romanian Journal on Information Science and Technology, Dan Tufiș (ed.) Special Issue on BalkaNet, Romanian Academy, vol7, no. 2-3, 2004, pp. 9-34, ISSN 1453-8245 (**68 citări**).  
*Prezintă principiile, metodologia și algoritmi de construcție asistată de calculator a unei ontologii lexical.*
6. **D. Tufiș**, Radu Ion, Nancy Ide: Fine-Grained Word Sense Disambiguation Based on Parallel Corpora, Word Alignment, Word Clustering and Aligned Wordnets. In Proc. of the 20<sup>th</sup> Intl. Conf. on Computational Linguistics, COLING2004, Geneva, 2004, pp. 1312-1318, ISBN 1-9324432-48-5 (**44 citări**).  
*Prezintă una dintre cele mai precise metode de dezambiguizare semantică a textelor, folosind cel mai fin inventar interligual de sensuri existent la ora actuală (Princeton WordNet).*
7. **D. Tufiș**: Tiered Tagging and Combined Classifiers, in F. Jelinek, E. Nöth (eds) Text, Speech and Dialogue, Lecture Notes in Artificial Intelligence 1692, Springer, 1999, pp. 28-33 (**39 citări**).  
*Prezintă modelul ierarhic de dezambiguizare morfo-lexicală larg utilizat în prezent de comunitatea științifică. O variantă mult extinsă a prezentării acestui model va apărea în monumentală lucrare "Oxford Handbook of Computational Linguistics" în curs de publicare la Oxford University Press.*
8. **D. Tufiș**: A Cheap and Fast Way to Build Useful Translation Lexicons. În Proc. of the 19th Intl. Conf. on Computational Linguistics, Taipei, 2002, pp. 1030-1036 (**35 citări**).  
*Introduce tehnologia extragerii automate de dicționare bilingve din texte paralele prin metode statistice. Reprezintă punctul de plecare al cercetărilor ulterioare în domeniul alinierii lexicale a textelor paralele.*
9. **D. Tufiș**: Using a Large Set of Eagles-compliant Morpho-Syntactic Descriptors as a Tagset for Probabilistic Tagging. Proc. of 2<sup>nd</sup> Intl. Conf. LREC2000, Athens 2000, pp. 1105-1112 (**29 citări**).  
*Descrie principiile fundamentale ale metodei dezambiguizării morfo-sintactice ierarhizate.*
10. **D. Tufiș**, Radu Ion, Alexandru Ceaușu, Dan Ștefănescu: Improved Lexical Alignment by Combining Multiple Reified Alignments. In Proc. of the 11<sup>th</sup> Conf. of the European Chapter of the Association for Computational Linguistics (EACL2006), Trento, Italy, 2006, pp. 153-160, ISBN 1-9324-32-61-2 (**29 citări**).  
*Descrie formalizarea modelului reificat de aliniere lexicală folosit în sistemul câștigător al competiției internaționale ACL2005, Ann Arbor, Michigan, USA..*
11. **D. Tufiș**, R. Ion, A. Ceaușu, D. Ștefănescu: RACAI's Linguistic Web Services. In Proc. of the 6<sup>th</sup> Language Resources and Evaluation Conference - LREC 2008, Marrakech, Morocco, May 2008. ELRA - European Language Resources Association (**25 citări**)  
*Descrie platforma web de prelucrări lingvistice multilingve automate folosită de peste 11780 de utilizatori din 1071 de orașe din 94 de țări. (conform Google Analytics Statistics).*
12. **D. Tufiș**, D. Cristea: Methodological issues in building the Romanian Wordnet and consistency checks in Balkanet. In Proc. of LREC2002 Workshop on Wordnet Structures and Standardisation, Las Palmas, Spain, 28-31 May, 2002, pp. 35-41 (**24 de citări**).  
*Prezintă principiile originale și metodologia lor de implementare în construcția ontologiilor lexicale aliniate bilingv.*

<sup>1</sup> O listă completă a publicațiilor de după 1997 și citările în baze de date internaționale publice (ISI, Google Scholar, ACM, CiteSeer, ACL Anthology) pot fi descărcate din pagina web personală: <http://www.racai.ro/~tufis>