

CANDIDAT PENTRU ALEGEREA CA MEMBRU TITULAR

Numele candidatului: Tufiş 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. Tufiş și-a început cercetările în domeniul inteligenței artificiale și a prelucrării automate a limbajului natural în 1980, ajungând prin realizările sale o personalitate științifică recunoscută pe plan mondial într-un domeniu de mare importanță atât națională cât și internațională. Eficacitatea sa în domeniu se explică prin urmărirea cu tenacitate a **ciclului complet al cercetării informatice**: de la formularea, analiza conceptuală și modelarea problemelor; la definirea de noi concepte, metode și algoritmi; la implementarea, testarea, evaluarea performanțelor și diseminarea rezultatelor. Aceste cercetări au condus la 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 200 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.). Google Scholar furnizează o listă de 1510 referințe, iar primele 20 dintre lucrările astfel identificate sunt referite de peste 600 de ori. A primit Premiul Academiei Române în 1989 și 1994. Prof. Tufiş a creat o școală românească de prestigiu internațional în domeniul prelucrării automate a limbajului natural, mulți dintre specialiștii pregătiți de el lucrând în universități și institute importante din lume. A înființat Comisia de Informatizare pentru Limba Română al cărui președinte este din anul 2001, for național cu cercetători din țară și din străinătate.

Priorități (listă parțială)

GENDATA și OFC: Printre primele sisteme software românești exportate (1980-1982), în R.D. Germană, Ungaria, URSS.

IURES2, PC-IURES, TC-LISP: Primele sisteme românești de Inteligență Artificială exportate. IURES2, omologat internațional în 1988, a introdus metode originale de prelucrare a limbajului natural. PC-IURES le-a perfecționat, devenind un sistem de referință în prelucrarea limbajului natural, recunoscut în 1992 de Japan Electronic Industry Development Association. TC-LISP, mediu de programare funcțională, a fost omologat internațional în anul 1988 și exportat în URSS, fiind considerat una dintre cele mai puternice implementări din lume pentru minicalculatoare. Monografia descriind TC-LISP și unele dintre sistemele implementate pe baza sa (IURES2, Discipol, Quernal), a apărut în 1987 la Editura Tehnică și a fost multă vreme cartea de referință pentru cercetătorii români din Inteligența Artificială.

Morfologia Paradigmatică și PARADIGM: Prima modelare și prima implementare din lume a morfologiei paradigmatică a unei limbi naturale, bazate pe unificare și structuri de tip atribut-valoare. Lucrarea "It Would Be Much Easier if Went Were Goed", a primit cel mai mare scor la conferința Europeană a Asociației de Lingvistică Computațională de la Manchester, UK, 1989.


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 și CLEF 2007 cu sisteme de întrebare-răspuns în spațiul web (întrebarea în română, răspunsul în documente în română sau engleză), întrecând sisteme din Spania, UK și România.

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. RO-WordNet este, după Princeton WordNet, cea mai mare și mai completă ontologie lexicală dintre cele 50 existente în lume, conținând definiții și axiome, marcaje de polaritate subiectiv-obiectiv și pozitiv-negativ pentru peste 54,000 de concepte, mai mult de 80,000 de relații semantice și lexicale și peste 80,000 de cuvinte.

În perioada 1990-2008, prin rezultatele sale excepționale, D. Tufiş a câștigat 34 de proiecte internaționale de cercetare și 10 naționale, ce au susținut cercetarea românească cu aproape 3.000.000 euro.



2 iunie, 2009

Acad. Mihai Drăgănescu, 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 RĂSUNET INTERNAȚIONAL (DE LA PRIMIREA ÎN ACADEMIA ROMÂNĂ)¹

1. D. Tufiș, P. Andersen (eds.) **“Recent Advances in Romanian Language Technology”**, Editura Academiei, 1997, 277p. (publicată și pe Internet cu sprijinul CE) *Volumul a fost considerat aproape doi ani (1998-1999) cea mai bine cotate publicație web, în domeniul prelucrării limbajului natural, de către situl de informații academice INFOSEEK (SUA) la două categorii de publicații.*
2. D. Tufiș, O. Mason: **Tagging Romanian Texts: A Case Study for QTAG, a Language Independent Probabilistic Tagger**, First Intl. Conf. on Language Resources and Evaluation, Granada, May, 1998, pp. 589-596, Prezintă unul dintre primele dezambiguatoare morfo-lexicale automate, bazat pe un model pur statistic. (**ACL Anthology**²)
3. 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. *Prezintă modelul ierarhic de dezambiguizare morfo-lexicală larg utilizat în prezent de comunitatea științifică.* (**ISI proceedings**)
4. 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. *A introdus tehnologia extragerii automate de dicționare bilingve din texte paralele prin metode statistice. A constituit punctul de plecare a cercetărilor ulterioare în domeniul alinierii lexicale a textelor paralele.* (**ACL Anthology**)
5. D. Tufiș, A.M. Barbu, R. Ion: **„TREQ-AL: A Word-alignment System with Limited Language Resources”**, Proc. of the North American Association of Computational Linguistics (NAACL) Workshop on Building and Using Parallel Texts; Romanian-English Shared Task, Edmonton, Canada, 2003, pp. 36-39. *Prezintă sistemul TREQ-AL, câștigător al competiției internaționale de aliniere lexicală pentru limbile română-engleză, organizată de NAACL, Edmonton, Canada 2003.* (**ACL Anthology**)
6. D. Tufiș, R. Ion, N. Ide: **Fine-Grained Word Sense Disambiguation Based on Parallel Corpora, Word Alignment, Word Clustering and Aligned Wordnets**. Proc. of the 20th Intl Conf. on Computational Linguistics, Geneva, 2004, pp.1312-1318. *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).* (**ACL Anthology**)
7. D. Tufiș, A.M. Barbu, R. Ion, **Extracting Multilingual Lexicons from Parallel Corpora**, Computers and the Humanities, Volume 38, Issue 2, 2004, pp. 163–189. *Prezintă metode, algoritmi și rezultate noi în construcția automată de dicționare multilingve din texte paralele (traduceri reciproce)* (**ISI journal**)
8. D. Tufiș, R. Ion, A. Ceașu, D. Ștefănescu: **„Combined Aligners”**. Proc. of the ACL2005 Workshop on “Building and Using Parallel Corpora: Data-driven Machine Translation and Beyond”, Ann Arbor, USA, June 2005, pp. 107-110, *Prezintă sistemul COWAL, câștigător al competiției internaționale de aliniere lexicală pentru limbile română-engleză, organizată de ACL, Ann Arbor, MI, SUA 2005.* (**ACL Anthology**)
9. D. Tufiș, R. Ion, A.Ceașu, D. Ștefănescu: **“Improved Lexical Alignment by Combining Multiple Reified Alignments”**. Proc. of the 11th Conf. of the European Chapter of the Association for Computational Linguistics (EACL 2006), Trento, Italy, April 2006, pp.153-160. *Prezintă un nou model de aliniere lexicală, modelul reificat al echivalențelor de traducere care a îmbunătățit substanțial performanțele sistemului COWAL.* (**ACL Anthology**)
10. D. Tufiș: **“Exploiting Aligned Parallel Corpora in Multilingual Studies and Applications”**. T. Ishida, S. R. Fussell and P. Vossen (eds.): “Intercultural Collaboration. 1st Intl Workshop, IWIC 2007 Kyoto, Japan, 2007. Invited and Selected Papers”. LNCS 4568, Springer, 2007, pp. 103-117. *Introduce metode colaborative noi pentru cercetarea multilinguală și multiculturală clădite pe teoria proprie a aliniierilor reificate ale textelor paralele.* (**ISI proceedings**)
11. D. Tufiș: **“Paradigmatic Morphology and Subjectivity Mark-Up in the RoWordNet Lexical Ontology”** In H.N. Teodorescu, J. Watada, and L. Jain (eds.), Intelligent Systems and Technologies - Methods and Applications, Studies in Computational Intelligence, Springer, 2009, pp. 161–179. *Reprezintă cea mai recentă descriere a stadiului monumentalului proiect (început în 2001) de ontologie lexicală pentru limba română, aliniată cu ontologia limbii engleze.* (**ISI proceedings**)
12. D. Tufiș, D. Ștefănescu, R. Ion, A. Ceașu. **RACAI’s Question Answering System at QA@CLEF 2007**. In C. Peters et al. (Eds.), Lecture Notes in Computer Science, LNCS 5152, Springer 2008, pp. 3284-3291. *Prezintă un nou sistem deschis de întrebare-răspuns pentru limba română, câștigător detașat al competiției europene Cross Lingual Evaluation Forum (CLEF 2007)* (**ISI proceedings**)

¹ O listă completă a publicațiilor de după 1997 (inclusiv formatul PDF al acestora) precum și indexările în baze de date internaționale publice (Google Scholar, ACM, CiteSeer) și citările pot fi descărcate din pagina web personală: <http://www.racai.ro/~tufis>

² ACL Anthology, cea mai prestigioasă arhivă de publicații ale domeniului lingvisticii computaționale, furnizează statistici (citări, indice Hirsch, etc) referitoare la cele aproape 12,000 de lucrări și autorii lor (7726) clasificați după indicele Hirsch. Această antologie glosează doar doi cercetători români din țară: D. Tufiș (poziția 276, între primii 3,6% autori) și D. Cristea (poziția 470).