

## **EREDMÉNYEK ÖSSZEFOGLALÁSA**

A T 037821 pályázat keretében az eredetileg kitűzött munkaterv megvalósult. A részjelentésekben foglaltakon túlmenően az elért eredmények összefoglalása a következő:

### **1) Az információvisszakereső módszerek egységes formális leírása.**

Megadtuk a Boole-féle, a vektortér, a valószínűségi módszerek egységes tárgyalási keretét. Megadtuk a kapcsolat alapú Web-visszakereső módszerek egységes formális keretét. Az eredményeket az 1.-4 rangos publikációk formájában ismertettük.

### **2) Az információvisszakeresés kapcsolata más tudományágakkal**

Új kapcsolatokra mutattunk rá az információvisszakeresés és a következő területek között:

- információelmélet: diszkrimináló képesség mérése felhasználva a Shannon-féle információmennyiség fogalmát,
- számelmélet: a Webgráf fokszámeloszlása aranyametszés alapú,
- nyelvtechnológia: megmutatuk, hogy a magyar nyelvre általában nem érvényes a Zipf-törvény, és hogy a magyar nyelv „kisvilág” gráf szerkezetű,
- orvostudomány: kapcsolat alapú visszakereső módszert alkalmaztunk CT képek keresésére,
- bonyolultságelmélet: mind elméletileg, mind kísérletileg megvizsgáltuk az asszociatív módszer számítási bonyolultságát, és megmutattuk, hogy az polinom rendű,
- logika: kapcsolatokra mutattunk rá fuzzy és valószínűségi logikával.

Az elért eredményeket 1z 5.-13. rangos publikációk tartalmazzák.

### **3) Visszakeresés hatékonyságának mérése**

Megmutattuk (tesztadatbázison történő szabványos laboratóriumi mérés segítségével), hogy az asszociatív visszakereső módszer átlagos hatékonysága 0,6 (ez az érték 20%-kal haladja meg a klasszikus módszerek hatékonyságát). A megbízhatóság matematikai elméleten alapuló módszert adtunk meg Webkeresőmotor hatékonyságának mérésére, és azt alkalmaztuk magyar és amerikai keresők mérésére akronimok tekintetében. Megállapítottuk, hogy a magyar akronimok általában kevésbé azonosítják honlapjaikat, mint az angol akronimok. Felhasználók bevonásával in vivo mértük a NeuRadIR orvosi kereső rendszer pontosságát (0,5-0,7 közöttinek találtuk). Entrópia alapú indexkifejezés-kiválasztó eljárást adtunk meg, és megmutattuk, hogy ilyen módon a vektortér visszakereső módszer hatékonysága növelhető. Eredményeinket a 14.-19. rangos publikációk tartalmazzák.

### **4) Keresők fejlesztése**

Kifejlesztettük a saját, korábban kidolgozott, fúzióként asszociatív módszert használó, i2rMeta metakeresőmotort ([www.dcs.vein.hu/CIR](http://www.dcs.vein.hu/CIR) (MetaSearch)), valamint a NeuRadIR ([www.dcs.vein.hu/CIR](http://www.dcs.vein.hu/CIR) (NeuRadIR)) orvosi kereső rendszereket. Eredményeinket a 20.-24. rangos publikációk tartalmazzák.

## 5) Tesztadatbázisok fejlesztése

Kifejlesztettünk egy angol nyelvű orvosi tesztadatbázist, ennek segítségével mértük a NeuRadIR rendszer hatékonyságát. Kifejlesztettünk hat magyar nyelvű tesztadatbázist, ezeket a kisvilág jelenség és az asszociatív módszer vizsgálatában használtuk fel. Eredményeinket a 25.-27. rangos publikációk tartalmazzák.

## 6) Eredményeink hozzáférhetővé tétele

Eredményeink tananyag részeivé váltak a Pannon Egyetem Műszaki Informatikai Karán (B.Sc és Ph.D. képzésben), a megfelelő jegyzetek a hallgatók számára (de bármely érdeklődő számára is) ingyenesen elérhetőek a [www.dcs.vein.hu/CIR](http://www.dcs.vein.hu/CIR) (Lecture Notes) címen. A Pannon Egyetemen kívül az eredmények tananyag részeit képezik a következő egyetemeken is: Joint Advanced Student School München, Germany; University of Colorado at Denver, USA; Eidgenössische Technische Hochschule Zürich, Schweiz; University for Health Sciences, Medical Informatics and Technology, Austria; University of Buffalo The State University of New York, USA; Pannon Egyetem; Széchenyi Egyetem.

## PUBLIKÁCIÓK

1. Dominich, S. (2002). Paradox-free formal foundation of vector space model. *Proceedings of the Annual International Conference on Research and Development in Information Retrieval of the Association for Computing Machinery Special Interest Group for Information Retrieval ACM SIGIR 3<sup>rd</sup> International Workshop Mathematical/Formal Methods in Information Retrieval MF/IR '02*. Association for Computing Machinery ACM, University of Tampere, Tampere, Finland, 12-15 August, pp: 43-48.
2. Dominich, S. (2003). PageRank: quantitative model of Interaction Information Retrieval. *Proceedings of the 12<sup>th</sup> International World Wide Web Conference WWW '03 International Workshop on Mobile Web Technologies WF7*. World Wide Web Consortium WWW-C, Institute for Electrical and Electronics Engineers IEEE, John von Neumann Society for Computer Science, May 20-24, Budapest, pp:13-18.
- 3 Dominich, S., and Skrop, A. (2005). PageRank and interaction information retrieval. *Journal of the American Society for Information Science and Technology*. John Wiley & Sons, vol. 56, no. 1, pp: 63-69, ISSN 1532-2882, **SCI IF=2,086**
- 4 Dominich, S., Tuza, Zs., and Skrop, A. (2006). Formal Theory of Connectionist Web Retrieval. In: Herrera-Viedma, Pasi, G., and Crestani, F. (eds.) *Soft Computing in Web Information Retrieval*, Springer Verlag, pp: 161-194, ISBN: 3-540-31588-8
5. Crestani, F., Dominich, S., Lalmas, M., and van Rijsbergen, C.J. (2003). Advances in mathematical and formal methods in information retrieval. *Journal of the American Society for Information Science and Technology*. John Wiley & Sons, vol. 54, no. 4, pp: 281-285, ISSN 1532-2882, **SCI IF=2,086**

6. Dominich, S., Góth, J., Kiezer, T., and Szlávik, Z. (2004). Entropy-based interpretation of Retrieval Status Value-based Retrieval. *Journal of the American Society for Information Science and Technology*. John Wiley & Sons, vol. 55, no. 7, pp: 613-627, ISSN 1532-2882, **SCI IF=2,086**
7. Dominich, S., Góth, J., Horváth, M., and T. Kiezer (2005). 'Beauty' of the World Wide Web: Cause, Goal, or Principle. *Lecture Notes in Computer Science, LNCS 3408*, Springer, pp: 67-81, ISSN 0302-9743, **SCI IF=0,513**
8. Dominich, S., Kiezer, T. (2005). Zipf-törvény, kis világ és magyar nyelv. *Alkalmazott Nyelvtudomány*, 1-2 szám, pp: 5-24, ISSN 1587-1061
9. Dominich, S., Góth, J., Kiezer, T. (2006). NeuRadIR: neuroradiological retrieval system. *Computerised Medical Imaging & Graphics*, Elsevier, ISSN 0895-6111, **SCI JIF=1.09**.
10. Szabó, T., Góth, J., Dominich, S., Kozmann, Gy., Szolgay, P. (2003). Novel Neuroradiological Image Processing and Information Retrieval in a Telestroke System. *Health Data in the Information Society – Proceedings of the International Conference on Medical Informatics 2003*. IOS Press, Amsterdam-Berlin-Oxford-Tokyo, pp: 298-303, ISBN 0926-9630.
11. Dominich, S., and Tuza, Zs. (2003). Computational aspects of connectionist interaction information retrieval. *Proceedings of the Annual International Conference on Research and Development in Information Retrieval of the Association for Computing Machinery Special Interest Group for Information Retrieval ACM SIGIR 4th International Workshop Mathematical Formal Methods in Information Retrieval MF/IR '03*. Association for Computing Machinery ACM, University of Toronto, Toronto, Canada, pp: 96-104.
12. Dominich, S., and Horváth, M. (2004). 'Golden' properties of the World Wide Web. *Proceedings of the 27th Annual International Conference on Research and Development in Information Retrieval of the Association for Computing Machinery Special Interest Group for Information Retrieval ACM SIGIR International Workshop Mathematical Formal Methods in Information Retrieval MF/IR '04*. Association for Computing Machinery ACM, University of Sheffield, Sheffield, U.K., 25-29 July, 2004, pp: 75-90.
13. Dominich, S., Lalmas, M., and van Rijsbergen, C.J. (2005). Spaces, Logic, and Link Analysis in IR: Recent Advances From A Mathematical and Logical Perspective. *Information Retrieval*, Springer-Verlag, vol. 8, no. 2, pp: 175-179, ISSN 1386-4564, **SCI IF=1,231**
14. Dominich, S. (2003). Connectionist Interaction Information Retrieval. *Information Processing and Management*. Elsevier, vol 39, no 2, pp: 167-194, ISSN 0306-4573, **SCI IF=1,506**

15. Dominich, S., Góth, J., and Skrop, A. (2003). Usefulness of Institutions' Acronyms As Web Queries. *Lecture Notes in Computer Science, LNCS 2633*, Springer Verlag Berlin-Heidelberg, pp: 580-588, ISSN 0302-9743, SCI IF=0,513
16. Dominich, S., and Skrop, A. (2005). Measuring the identification capability of acronyms on the World Wide Web: a comparative study. *Journal of Web Engineering*, Rinton Press, vol, no. 3&4, pp: 220-215, ISSN 1540-9589
17. Dominich, S., Góth, J., Kiezer, T. (2006). NeuRadIR: neuroradiological retrieval system. *Computerised Medical Imaging & Graphics*, Elsevier, ISSN 0895-6111, SCI JIF=1.09.
18. Dominich, S, and Horváth, M. (2003) The application of Shannon information to the computation of concept discrimination value. *Proceedings of the International Conference on Information society, cultural heritage and text analysis*. University of Göteborg, Boras, Sweden, Technical University of Budapest, 24–26 November 2003, Budapest, Hungary, pp: 92-105, ISBN 963-421-584-X.
19. Dominich, S., Góth, J., Kiezer, T., and Szlávik, Z. (2004). Entropy-based interpretation of Retrieval Status Value-based Retrieval. *Journal of the American Society for Information Science and Technology*. John Wiley & Sons, vol. 55, no. 7, pp: 613-627, ISSN 1532-2882, **SCI IF=2,086**
20. Dominich, S. (2003). Connectionist Interaction Information Retrieval. *Information Processing and Management*. Elsevier, vol 39, no 2, pp: 167-194, ISSN 0306-4573, SCI IF=1,506
21. Dominich, S., Góth, J., Kiezer, T., and Szlávik, Z. (2003). NeuRadIR: neuroradiológiai információ-visszakereső rendszer. *Informatika és Menedzsment az Egészségügyben*. vol. 2. no 1, pp: 41-47, ISSN 1588-6387
22. Dominich, S., Góth, J., and Kiezer, T. (2005). NeuRadIR: A Web-Based NeuroRadiological Information Retrieval System. *ERCIM News*, ERCIM EEIG, No. 61, April, pp: 52-53, ISSN 0926-4981
23. Dominich, S., and Góth, J. (2002). Retrieval of Brain CT Reports and Images Using Interaction Information Retrieval. *Health Data in the Information Society – Proceedings of the International Conference on Medical Informatics 2002*. IOS Press, Amsterdam-Berlin-Oxford-Tokyo-Washington, 25-29 August, pp: 325-330, ISBN 1-58603-279-8
24. Szabó, T., Góth, J., Dominich, S., Kozmann, Gy., Szolgay, P. (2003). Novel Neuroradiological Image Processing and Information Retrieval in a Telestroke System. *Health Data in the Information Society – Proceedings of the International Conference on Medical Informatics 2003*. IOS Press, Amsterdam-Berlin-Oxford-Tokyo, pp: 298-303, ISBN 0926-9630.

25. Dominich, S., Góth, J., Kiezer, T. (2006). NeuRadIR: neuroradiological retrieval system. *Computerised Medical Imaging & Graphics*, Elsevier, ISSN 0895-6111, **SCI JIF=1.09**

26. Dominich, S., Kiezer, T. (2005). Zipf-törvény, kis világ és magyar nyelv. *Alkalmazott Nyelvtudomány*, 1-2 szám, pp: 5-24, ISSN 1587-1061

27. Dominich, S., Darányi, S., Szlávik, Z. (2006). Magyar hiedelmek hierarchikus struktúrája. *Alkalmazott Nyelvtudomány*, 1-2 szám, pp: 137-161, ISSN 1587-1061