

Richárd Farkas PhD

Nationality: Hungarian

Date of birth: 31st Jan 1980

homepage: <http://www.inf.u-szeged.hu/~rfarkas>

RESEARCH INTERESTS

Statistical Natural Language Processing (especially Information Extraction)

Novel models for language phenomena

Machine learning with “human-in-the-loop”

EDUCATION

- 2006 – 2009 **Computer Science PhD**, University of Szeged;
Thesis title: *„Machine Learning techniques for applied Information Extraction”*
- 2004 – 2007 **Economics MA**, University of Szeged
- 2001 – 2003 **Computer Science MSc**, University of Szeged
Thesis title: *„Genetic algorithms in optical lithography”*
supervised by the Fraunhofer Institute Erlangen, Germany.
- 1998 – 2001 **Computer Science BSc**, University of Szeged
Thesis title: *„Text Mining of enterprise quarterly reports”* (in Hungarian)

WORK EXPERIENCE

- 2010 – **Graduate research assistant** at the University of Szeged, Human Language Team, research topics: web mining, trend- and text-analysis
- 2004 – 2006 **Research engineer** at GusGus AB
research topic: machine learning algorithms for automatic trading on stock markets, supervisor: Prof. Csaba Szepesvári
- 2003 – 2004 **Graduate research fellow** at the Hungarian Academy of Sciences, Research Group on Artificial Intelligence, research topic: named entity recognition
- 2001 – 2003 **Software developer** at NPT Szoftver Bt.

RESEARCH PROJECTS

- 2009 "Automatic free-text-tagging of the [origo] online news archive", sold to the Origo news portal, role: principal research investigator
- 2009 – 2010 "Automatic Construction of chemo-, and biological databases", sponsored by the Hungarian Sponsored by the National Office for Research and Technology, role: natural language processing expert (principal investigator: Dr. Zsolt Molnár, Acheuron Ltd.)
- 2009–2010 "Criminal information extraction system", sponsored by the Hungarian Coordination Centre Against Organized Crime, role: principal research investigator
- 2008–2010 "TEXTREND: Development of a business and governmental decision support toolbox using trend- and text-analysis tools", Sponsored by the National Office for Research and Technology, role: researcher (principal investigator: Dr. Márk Jelasity, Hungarian Academy of Sciences)
- 2008 – 2010 „BelAml: Ambient Intelligent Systems”, Sponsored by the National Office for Research and Technology, role: machine learning expert (principal investigator: Prof. János Csirik, University of Szeged)
- 2003 „Information Extraction from Short Business News”, Sponsored by the National Office for Research and Technology, role: researcher (principal investigator: Prof. János Csirik, University of Szeged)

AWARDS, PRIZES, SHOLARSHIPS

- 2009 Gábor Berend and István Nagy won I. and II. prize at the National Scientific Conference of Students (Hungary) under my supervision
- 2008 2nd place at the [Obesity Challenge](#) /clinical document labeling/ (team Szeged)
- 2007 1st place at [Classifying Clinical Free Text Using Natural Language Processing](#) (team Szeged)
- 2007 1st place at the [Metonymy resolution](#) shared task (team GYDER)
- 2006 2nd place at [the De-identification of medical records](#) /clinical Named Entity Recognition/ shared task (team Szeged)
- 2006 best student work award at the Hungarian Computational Linguistics conference
- 2004 Advanced Retrieval and Web Mining, PhD summer school, University of Bologna, Italy
- 2003 Best Industrial application prize at the National Scientific Conference of Students (Hungary)

- 2003 6 months ERASMUS scholarship in Germany, Friedrich Alexander
Universität Erlangen-Nürnberg
- 1998 2nd place at the László Kalmár software developer contest, Hungary

TEACHING EXPERIENCE

- 2007 – 2010 Human Language Technologies (newly designed graduate and
undergraduate course)
- 2007 – 2009 Artificial Intelligence (undergraduate course)
- 2006 – 2008 Basic Algorithms (undergraduate course)
- 2003 – 2004 Software design (undergraduate course)

REVIEWING

PROGRAM COMMITTEE MEMBER

- 2010 Computational Natural Language Learning (CoNLL) [Shared Task](#),
head of the shared task organising team
- 2010 [Louhi Workshop](#) on Text and Data Mining of Health Documents
(a NAACL workshop)
- 2009 [Empirical Methods in Natural Language Processing](#) (EMNLP)
- 2009 [International Journal of Medical Informatics](#)

SUPERVISED MSc STUDENTS

- | | | |
|------|-----------------|---|
| 2010 | Gábor Molnár | Multi-layer Named Entity Recognition |
| 2009 | Gábor Berend | Opinion Mining based on textual and graphical clues |
| 2009 | István Nagy | Web mining of enterprise contact information |
| 2009 | György Móra | Biological event extraction from texts |
| 2009 | Krisztina Szabó | Topic identification of movies based on subtitles |
| 2009 | Orsolya Németh | Parsing of medical texts |
| 2008 | Péter Ótott | Biological Named Entity Recognition |

PROGRAMMING SKILLS

- Java 4 years of programming experience
- C++ 5 years of programming experience
- shell scripts 5 years of programming experience

LANGUAGES SPOKEN

- Hungarian (mother tongue)
- English (fluent)
- German (fluent)

SELECTED PUBLICATIONS¹

Richárd Farkas, Berend Gábor, István Hegedűs, András Kárpáti, Balázs Krich: **Automatic free-text-tagging of online news archives**, accepted to the 19th European Conference on Artificial Intelligence - ECAI 2010. (nominated for the best paper award.)

Richárd Farkas, Veronika Vincze, György Móra, János Csirik, György Szarvas: **The CoNLL 2010 Shared Task: Learning to Detect Hedges and their Scope in Natural Language Texts**, In Proceedings of the Fourteenth Conference on Computational Natural Language Learning (CoNLL 2010): Shared Task, pp 1-12

István Nagy, Richárd Farkas, Márk Jelasity: **Researcher affiliation extraction from homepages**, in Proceedings of the NLP4DL ACL Workshop 2009 pp 1-9

Richárd Farkas: **The strength of co-authorship in gene name disambiguation**, BMC Bioinformatics 2008, 9:69

Filip Ginter, Richárd Farkas: **Automated assignment of ICD-9-CM codes to radiology reports**. Invited talk at The First Louhi Conference on Text and Data Mining of Clinical Documents, 2008

Richárd Farkas and György Szarvas: **Automatic construction of rule-based ICD-9-CM coding systems**, BMC Bioinformatics 2008

György Szarvas, Richárd Farkas, Róbert Busa-Fekete: **State-of-the-art anonymisation of medical records using an iterative machine learning framework**. Journal of the American Medical Informatics Association, 2007 Vol. 14. pp 574-580

György Szarvas, Richárd Farkas, András Kocsor: **A Multilingual Named Entity Recognition System Using Boosting and C4.5 Decision Tree Learning Algorithms**. The Ninth International Conference on Discovery Science 2006, LNAI 4265

Farkas Richárd: **Tulajdonnév-felismerés** In Tikk Domonkos (editor): Szövegbányászat, 2007 TypoTex (bookchapter in Hungarian, title: Named Entity Recognition)

¹ For a full list of publications please visit: <http://www.inf.u-szeged.hu/~rfarkas/publications.html>