

# Richárd Farkas PhD

---

Nationality: Hungarian

Date of birth: 31. Jan 1980

homepage: [www.inf.u-szeged.hu/~rfarkas](http://www.inf.u-szeged.hu/~rfarkas)

## RESEARCH INTERESTS

Statistical Natural Language Processing (especially Information Extraction),  
Structured prediction and discriminative ranking problems,  
Exploiting extra-textual knowledge in Natural Language Processing,  
Linguistic processing of morphologically rich languages.

## WORK EXPERIENCE

- 2013 - assistant professor at the University of Szeged,  
research topics: deep analysis of (user-generated) texts
- 2011 - 2012 postdoc at the IMS Stuttgart Statistical Natural Language Processing Group,  
research topic: discriminative parsing for morphologically rich languages and non-  
canonical texts  
supervisors: Hinrich Schütze and Helmut Schmid
- 2009 - 2010 graduate teaching assistant at the University of Szeged,  
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 (reinforcement learning, time series prediction),  
supervisor: Csaba Szepesvári
- 2003 - 2004 graduate research fellow at the Hungarian Academy of Sciences, Research  
Group on Artificial Intelligence,  
research topic: named entity recognition

## EDUCATION

- 2006 - 2009 Computer Science PhD, University of Szeged;  
Thesis title: „*Machine Learning techniques for applied Information Extraction*”
- 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)

## AWARDS, PRIZES, SHOLARSHIPS

2014	“Farkas Gyula Prize” of the János Bolyai Mathematical Society (for applications of mathematics)
2013-2014	Zoltán Magyary scholarship, Hungarian National Excellence Program
2009	Gábor Berend and István Nagy won I. and II. prize at the National Scientific Conference of Students (Hungary) under my supervision
2007	1 <sup>st</sup> place at <a href="#">Classifying Clinical Free Text Using Natural Language Processing</a> (team Szeged)
2007	1 <sup>st</sup> place at the <a href="#">Metonymy resolution</a> shared task (team GYDER)
2006	2 <sup>nd</sup> place at <a href="#">the De-identification of medical records</a> /clinical Named Entity Recognition/ shared task (team Szeged)
2006	best student work award at the Hungarian Computational Linguistics conference
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	2 <sup>nd</sup> place at the László Kalmár software developer contest, Hungary

## SELECTED PAST RESEARCH PROJECTS

As principal investigator:

2012 – 2014	“Infocommunicational technologies and the society of the future”, role: leader of the natural language processing subproject, team size: 14
2013 – 2014	“Intelligent recruitment systems”, sponsored by the Nexum Ltd., role: principal investigator, team size: 5
2010 – 2013	“Criminal information extraction system”, sponsored by the Hungarian Coordination Centre Against Organized Crime, team size: 2
2010	“Automatic keyphrase extraction from Hungarian news”, sponsored by the [origo] news portal, role: principal investigator, team size: 3

As a researcher:

2011 – 2012	“ <a href="#">Incremental Specification in Context (SFB 732)</a> ”, role: postdoc researcher (at the Institute for Natural Language Processing, University of Stuttgart)
2009 – 2010	“Automatic Construction of chemo-, and biological databases from text”, role: natural language processing expert (principal investigator: Dr. Zsolt Molnár, Acheuron Ltd.)
2008 – 2010	“ <a href="#">TEXTREND</a> : Development of a business and governmental decision support toolbox using trend- and text-analysis tools“, role: researcher (principal investigator: Dr. Márk Jelasity, Hungarian Academy of Sciences)

## PUBLICATIONS

91 publications (88 in the field of NLP), 63 in English (28 in Hungarian), 30 first author  
12 journal articles, 1 book chapter, 1 Phd thesis, 77 conference papers  
impact factor=21.3, independent citations>400, h-index=12, Erdős-number=3

## SELECTED PUBLICATIONS<sup>1</sup>

Zsolt Szántó, Richárd Farkas: **Special Techniques for Constituent Parsing of Morphologically Rich Languages**. In Proceedings of EACL 2014 pp 135-144.

Alexander Fraser, Helmut Schmid, Richárd Farkas, Renjing Wang and Hinrich Schuetze: **Knowledge Sources for Constituent Parsing of German, a Morphologically Rich and Less-Configurational Language**. Computational Linguistics, Volume 39, Issue 1. 2013.

Richárd Farkas, Helmut Schmid: **Forest Reranking through Subtree Ranking**. in the Proceedings of Conference on Empirical Methods in Natural Language Processing EMNLP-2012. pp 1038-1047

Richárd Farkas, Veronika Vincze, Helmut Schmid: **Dependency Parsing of Hungarian: Baseline Results and Challenges**. in the Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics EACL-2012, pp 55-65

György Szarvas, Veronika Vincze, Richárd Farkas, György Móra, Iryna Gurevych: **Cross-Genre and Cross-Domain Detection of Semantic Uncertainty**. Computational Linguistics, Volume 38, Issue 2, Special Issue on Modality and Negation. 2012.

Richárd Farkas: **Learning Local Content Shift Detectors from Document-level Information**. In Proceedings of the Conference on Empirical Methods in Natural Language Processing – EMNLP 2011 pp 759-770

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

Richárd Farkas, Berend Gábor, István Hegedűs, András Kárpáti, Balázs Krich: **Automatic free-text-tagging of online news archives**, in the Proceedings of the 19th European Conference on Artificial Intelligence pp 529-534

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

---

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

## **SUPERVISION OF STUDENTS**

- 2010- 4 PhD students at the University of Szeged, Doctoral School of Informatics (1 after defence, +1 submitted thesis)
- 2013- 3 MSc and 2 BSc theses at the University of Szeged, Institute of Informatics
- 2011 1 MSc thesis at IMS Stuttgart, Germany
- 2006 – 2010 9 MSc and 2 BSc theses at the University of Szeged, Institute of Informatics (1st and 2nd prizes at the Hungarian National Scientific Conference of Students)

## **TEACHING ACTIVITIES** (all at the University of Szeged, Institute of Informatics)

- 2014 - Intelligent Human-Computer Interactions – lectures (graduate course)
- 2013 – Introduction to Machine Learning – lectures (mandatory graduate course)
- 2008 – Natural Language Processing – lectures (new optional graduate course)
- 2007 – 2010 Artificial Intelligence – exercises (undergraduate course)

## **ORGANISATION OF SCIENTIFIC MEETINGS**

- 2010 Head of the organising team of the [Conference on Natural Language Learning 2010 Shared Task](#)
- 2013 – to date Member of the organizing committee of the Hungarian Computational Linguistics annual conference

## **DATASETS AND CODES** (each of them is publicly available)

- [magyarlanc](#) Toolchain for linguistic preprocessing of Hungarian
- [CSD](#) Local content shift detection algorithms
- [NER](#) Named Entity Recognition toolkit enabling simple feature engineering
- [BioScope corpus](#) Biological text annotated for negation and uncertainty along with their scope
- [forum corpus](#) Hungarian forum entries annotated for Opinion Mining
- [NElemma](#) Named Entity lemmatization database (Hungarian and English)

## **REVIEWING ACTIVITY**

CONFERENCE PAPERS AT:

EACL, LREC, IJCNLP, EMNLP, SPMRL, CoNLL Shared Task, Louhi

JOURNAL ARTICLES IN:

Special Issue of the Computational Linguistics Journal on Modality and Negation,

Journal of the American Medical Informatics Association,

Knowledge and Information Systems,

International Journal of Medical Informatics,

Általános Nyelvészeti Tanulmányok, Acta Cybernetica

## **LANGUAGES SPOKEN**

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