

RESEARCH TEAM PROFILES

FACULTY

Faculty of Business and Economics

RESEARCH AREA

Text Mining, Natural Language Processing, Machine Learning, Sentiment Analysis, Plagiarism Detection

RESEARCH TEAM

Assoc. Prof. František Dařena, team leader, contact: frantisek.darena@mendelu.cz, tel.: +420 545 132 229

Assoc. Prof. Jan Žižka, Mendel University in Brno, Czech Republic

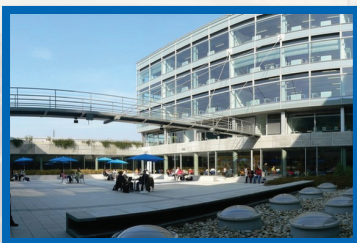
Jan Přichystal, PhD., Mendel University in Brno, Czech Republic

Tomáš Foltýnek, PhD., Mendel University in Brno, Czech Republic

Jiří Krupník, Mendel University in Brno, Czech Republic

EXPERT/TECHNICAL PROFILE OF THE TEAM (SPECIFIC PART OF RESEARCH)

The Natural Language Processing and Text Mining Group at the Faculty of Business and Economics, Mendel University in Brno concerns with semantic processing of electronic text data written in natural languages with the goal to reveal knowledge hidden in the data. Various tools for data pre-processing, including stemming, stop word removal, spell checking and others are used and being developed. For the transformation of the data to a format suitable for machine learning algorithms and different software tools, a unique application (with optional graphical user interface) is being continuously developed. Advanced supervised (classification), unsupervised (clustering, feature selection, associations) and semi-supervised machine learning techniques and their combinations are applied to texts. The goals of the exploration include categorization of text documents, retrieval of documents based on similarity, discovering the semantics of groups of documents, finding attributes that express meaning, sentiment analysis, and others. Specific conditions and limitations are taken into consideration, for example, low number of good examples, huge volumes of data, curse of dimensionality, sparseness of vectors representing the data, imbalanced classes, or multilingualism, which is – unlike usual scientific papers or newspaper articles – typical for collections of data generated by users on social networks, microblogging sites, or discussion boards. The future work is also focused on analysing relations between the text data (news, economic reports, social network posts and likes) and various economic phenomena, such as stock price development.



EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND
INVESTING IN YOUR FUTURE



EXPERIENCE

Participation in national projects (programme and provider)

- Czech Economics in Integration and Globalization Processes and the Agrarian Sector Development within the Integrated European Market Conditions (Czech ministry of Education, 2005–2011)
- Analysis of opinions hidden in large unstructured text data written in natural language with application of data mining methods (Internal Grant Agency, Mendel University, 2012)
- Analysis of text data pre-processing impact on text mining tasks (Internal Grant Agency, Mendel University, 2012)
- Automated support for searching for semantic information and knowledge in text data (Internal Grant Agency, Mendel University, 2012)

Conference organization etc. (national and international)

- The Third International Workshop SoNet–2012, September 17–21, 2012 in conjunction with the 5th International Conference on Intelligent Information and Engineering Systems INFOS-2012, Rzeszow, Poland
- The Second International Workshop on Social Networks: Computing and Mining – SoNet-2010, September 3–5, 2010, Brno, Czech Republic; organized by SoNet Research Center, Department of Informatics, Mendel University in Brno, Czech Republic, fLexSem Laboratory, Department of French and Romance Philology Autonomous University of Barcelona, Spain, and Statistical Cybermetrics Research Group, the School of Computing and Information Technology at the University of Wolverhampton, Wolverhampton, United Kingdom
- The First International Workshop on Social Networks and Their Applications Tools September 19–21, 2008, University of Central Europe (UCE), Královská 386/11, 909 01 Skalica, Slovakia
- International conference Plagiarism across Europe and beyond, June 12–14, 2013
- International conference Plagiarism across Europe and beyond, June 10–12, 2015

Involvement in European (especially in Framework Programmes projects, EUREKA, EUROSTARS, COST etc.) and international cooperation

- Partner of the LLP/Erasmus project Impact of Policies for Plagiarism in Higher Education across Europe (coordinated by Coventry University, UK)

Other foreign activities and contacts (active participation in conferences, publications, membership in international bodies etc.)

Team members are regular members of committees at many international conferences (e.g., 3rd International Conference on Multimedia Computing and Systems 2012, IADIS International Conference WWW/INTERNET 2012, 2013, 2014, 3rd International Conference on Advances in Information Mining and Management 2013, 4th International Conference on Multimedia Computing and Systems 2014, 13th International Conference WWW/INTERNET 2014, 4th International Conference on Advances in Information Mining and Management 2014), reviewers of international journals (e.g., Knowledge Based Systems, Information and Software Technology), participants in international conferences and workshops, teachers at foreign universities (e.g., Saint Petersburg State University and the Russian Presidential Academy of National Economy and Public Administration, Moscow, Russia, Department of French and Romance Philology, Faculty of Philosophy and Arts, Autonomous University of Barcelona, Spain, University of Debrecen, Hungary, Universitat Politècnica de València, Spain), editors-in-chief or editors of international journals (e.g., International Journal in Foundations of Computer Science & Technology, International Journal of Intelligent Information Processing, PICA: Progress in Intelligent Computing and Applications, International Journal of Information Sciences and Techniques, International Journal of Computer Science & Information Technology).

Topics for cooperation in H2020/project proposal

- semantic processing of electronic text data written in natural languages with the goal to reveal knowledge hidden in the data
- analysis of text data pre-processing on text mining tasks
- application of advanced machine learning algorithms – supervised (classification), unsupervised (clustering, feature selection, associations) and semi-supervised – and their combinations to text data
- categorization of text documents, retrieval of documents based on similarity, discovering the semantics of groups of documents, finding attributes that express meaning, sentiment analysis, and others
- solving problems with low number of good examples, huge volumes of data, curse of dimensionality, sparseness of vectors representing the data, imbalanced classes, or multilingualism
- analysis of relations between the text data (news, economic reports, social network posts and likes) and various economic phenomena, such as stock price development



EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND
INVESTING IN YOUR FUTURE

