Proposal for Cooperation of Intelligent Computer Systems Department of NTU “KhPI”

ABOUT US

- Intelligent Computer Systems Department (http://web.kpi.kharkov.ua/iks/en/)
- Bachelor and Master Programs “Computational Linguistics”

Core courses:

- Modern technologies of Internet-applications development;
- Ontology engineering;
- Intelligent data analysis;
- Modern software of text analysis;
- Corpus linguistics;
- Formal modeling of natural language objects;
- Artificial intelligence.

NATALIA SHARONOVA,
Doctor of Engineering, Full Professor

sharonova@kpi.kharkov.ua,

nvsharonova@mail.ru

RESEARCH TOPICS

- Web-Based Monitoring and Evaluation;
- Natural language processing;
- Text Mining and Web Content Mining;
- Processing of semi-structured textual data;
- Sentiment analysis and customer emotion score;
- Information extraction
  (name-entity extraction + relationship recognition);
- Special lexicons (Gazetteers);
- Vision & Language area.

OUR PROJECTS

- MASTIS - Establishing Modern Master-level Studies in Information System (Erasmus+, 2015-2017)
- AComIn - Advanced Computing for Innovation (Bulgaria, 2015-2016), FP7 Capacity Programme, Research Potential of Convergence Regions, Call: FP7-REGPOT-2012-2013-1
- One-stage semantic analysis of non-literary texts (Samsung, Ukraine, 2012)
- Models and methods development of Information Retrieval for economic topics (Switzerland, 2011)

OUR PUBLICATIONS

- Khairova, N.F.,Petrasova, S.,Gautam, A.P.S. The logical-linguistic model of fact extraction from English texts . Information and Software Technologies. Volume 639 of the series Communications in Computer and Information Science, , Springer, ISBN: 978-3-319-46253-0, 2016, pp. 625-635. doi> 10.1007/978-3-319-46254-7_51
- O. Kanishcheva, G. Angelova, S. G. Nikilov, Towards Translation of Tags in Large Annotated Image Collections, Artificial Intelligence: Methodology, Systems, and Applications, Volume 9883 of the series Lecture Notes in Computer Science, Springer, ISBN: 978-3-31944747-6, 2016, pp. 140-150. doi> 10.1007/978-3-319-44748-3_14
- Van-Hieu Vu, Hai-Son Le, O. Kanishcheva and G. Angelova, Fine-tuning SIMPLE based Content Based Image Retrieval system, Proceedings of the 6th International Symposium on Information and Communication Technology (SoICT 2015), 03-05 December 2015, Hue, Vietnam, ISBN: 978-1-4503-3843-1, Paper id: 36, 8 pages. doi>10.1145/2833258.2833273
- Kanishcheva O. and G. Angelova. *About Emotion Identification in Visual Sentiment Analysis*, In: G. Angelova, K. Bontcheva, R. Mitkov (Eds.), Proceedings of the 10th International Conference on "Recent Advances in Natural Language Processing" RANLP 2015, 7-9 September 2015, Hissar, Bulgaria, ISSN: 1313-8502, pp. 258-265. Available at http://lml.bas.bg/ranlp2015/docs/RANLP_main.pdf

OUR PUBLICATIONS (2)

- Khairova, N., Petrasova, S., Ajit Pratap Singh Gautam. The logical and linguistic model for automatic extraction of collocation similarity // Econtechmod : an international quarterly journal on economics in technology, new technologies and modelling processes. – Lublin; Rzeszow, 2015. – № 3 (4). – P. 43–48. Available at http://www.pan-ol.lublin.pl/wydawnictwa/Econtechmod2015/Econ_2015_4.pdf
- Kanishcheva, O. and G. Angelova. *A Pipeline Approach to Image Auto-Tagging Refinement*. In: Bădică C., Y. Manolopoulos, M. Coșulschi, G. Eleftherakis and F. Leon (Eds.) Proceedings of the 7th Balkan Conference on Informatics Conference, Craiova, Romania, ACM New York, NY, USA, 2015, ISBN: 978-1-4503-3335-1, Paper id: 9, 8 pages. doi>10.1145/2801081.2801108
- O. Kanishcheva, G. Angelova, About Sense Disambiguation of Image Tags in Large Annotated Image Collections, Innovative Approaches and Solutions in Advanced Intelligent Systems, Volume 648 of the series Studies in Computational Intelligence, Springer, ISBN: 978-3-319-32207-0, 2016, pp. 133-149. doi> 10.1007/978-3-319-32207-0_9
- Khairova N., Sharonova N. Building of Logic Network of the Information Area of Corporation // 8th IEEE EAST-WEST DESIGN&Test Symposium (EWDTS 2010) / St. Peterburg, Russia, September 17-20, 2010. – P. 371 – 373.
- Sharonova N. Use of Predicate Categories for Modelling of Operation of the Semantic Analyzer of the Linguistic Processor / N. Sharonova, N. Khairova // Proceeding of IEEE East-West Design & Test Symposium (EWDTS'09). – Moscow, 2009. – P. 204-208.

OUR PROPOSAL

PROJECT "THE AUTOMATED ASSESSMENT OF THE QUALITY OF SEMI-STRUCTURED TEXTUAL INFORMATION"

The purpose of this research is to develop and to experimentally evaluate technology for automated assessment of the quality of semi-structured textual information. The technology is based on mathematical and computational methods of natural language processing. Technical documentation, such as user manuals and technical specifications in several languages (English, Ukrainian, Russian, German, and others), is the main object for this research study and evaluation of the developed technology.

In order to succeed it's necessary to consider the existing structural, syntactic, lexical, and semantic characteristics of technical documentation in the specified languages. Afterwards, the multilingual thesaurus and semantic network for the analysis of such texts has to be elaborated based on these characteristic.

To model semantic relations between concepts in technical documentation we will use the elaborated semantic network and the mathematical theory of intelligence (algebra of finite predicates and description logics).

Since quality assessment of technical documentation is a feature of intelligence, semantic analysis of texts based on modelling of the text comprehension has to be applied.

Therefore, the approach of the semantic analysis, which models text comprehension of the documents using mathematical theory of intelligence, is proposed in this research. Finally, experimental evaluation of the created models for a specific technical documentation has to be performed.

OUR PROPOSAL

PROJECT "NAME ENTITY RECOGNITION AND OPINION IDENTIFICATION FOR TOURISM AREA"

The main goal is to develop software for extracting entities (such as sights, museums etc.) to create ontologies of tourist objects. Our models also will detect and disambiguate entity mentions and perform sentiment analysis and opinion detection on the texts, to be able for example, to extract the sentiment and the opinion of customers about certain sights in Web reviews, blogs, Twitter etc.

Our project will allow to process results from different Internet sources that will allow to create more relevant and various list of entities.

Project results can be used for various application areas and tailored for different languages, including the Ukrainian.