

|                         |  |
|-------------------------|--|
| 1. Record Nr.           | UNINA9910299291103321  |
| Titolo                  | Artificial Intelligence and Natural Language : 6th Conference, AINL 2017, St. Petersburg, Russia, September 20–23, 2017, Revised Selected Papers // edited by Andrey Filchenkov, Lidia Pivovarova, Jan Žižka   |
| Pubbl/distr/stampa      | Cham : , : Springer International Publishing : , : Imprint : Springer, , 2018  |
| ISBN                    | 3-319-71746-4  |
| Edizione                | [1st ed. 2018.]  |
| Descrizione fisica      | 1 online resource (XI, 305 p. 39 illus.)   |
| Collana                 | Communications in Computer and Information Science, , 1865-0929 ; ; 789  |
| Disciplina              | 006.3  |
| Soggetti                | Artificial intelligence<br>Information storage and retrieval<br>Natural language processing (Computer science)<br>Data mining<br>Optical data processing<br>Artificial Intelligence<br>Information Storage and Retrieval<br>Natural Language Processing (NLP)<br>Data Mining and Knowledge Discovery<br>Image Processing and Computer Vision   |
| Lingua di pubblicazione | Inglese  |
| Formato                 | Materiale a stampa   |
| Livello bibliografico   | Monografia   |
| Note generali           | Includes index.  |
| Nota di contenuto       | Intro -- Preface -- Organization -- Contents -- Social Interaction Analysis -- Semantic Feature Aggregation for Gender Identification in Russian Facebook -- 1 Introduction -- 2 Related Work -- 2.1 Feature Aggregation for Author Profiling in Social Media -- 2.2 Topic Modelling -- 2.3 Distributional Clustering -- 3 Dataset -- 4 Feature Aggregation Models -- 4.1 LDA -- 4.2 Author-Topic Model -- 4.3 Distributional Clustering -- 4.4 Automatic Label Assignment -- 5 Author Gender Profiling -- 5.1 Experiment -- 5.2 Results -- 5.3 Correlation Analysis -- 6 Conclusions -- References -- Using Linguistic Activity in Social Networks to Predict and Interpret Dark Psychological Traits -- 1 Introduction -- 2 Method -- 2.1 Psychometrics -- 2.2 Topic Models -- |

2.3 Predictive Models -- 2.4 Statistical Analysis -- 3 Experiment -- 3.1 Data Collection -- 3.2 Data Preprocessing -- 3.3 Implementation Details -- 4 Results -- 4.1 Prediction -- 4.2 Statistical Analysis -- 5 Discussion -- 6 Conclusion -- References -- Boosting a Rule-Based Chatbot Using Statistics and User Satisfaction Ratings -- 1 Introduction -- 2 Related Work -- 3 Overview of the Rule-Based Chatbot -- 4 Raw Data and Task Definition -- 4.1 Data -- 4.2 Approach Chosen -- 5 Data Preparation -- 6 Experimental Setup -- 6.1 Data Preprocessing -- 6.2 Baselines -- 6.3 Proposed Systems -- 7 Results and Discussion -- 7.1 System Performance -- 7.2 Difficulty of the Task -- 8 Conclusion and Future Work -- References -- Speech Processing -- Deep Learning for Acoustic Addressee Detection in Spoken Dialogue Systems -- Abstract -- 1 Introduction -- 2 Related Work -- 3 Experimental Data -- 4 Features -- 5 Models -- 6 Results -- 7 Conclusions -- Acknowledgments -- References -- Deep Neural Networks in Russian Speech Recognition -- 1 Introduction -- 2 Related Work -- 3 Architectures of Neural Networks for Acoustic Modeling. 3.1 LSTM -- 3.2 CNN -- 3.3 ResNet -- 3.4 RCNN -- 4 Datasets -- 4.1 Dataset for the Acoustic Models -- 4.2 Dataset for the Language Model -- 5 Speech Recognition System Implementation -- 6 Experiments and Results -- 6.1 Baseline -- 6.2 MLP -- 6.3 LSTM -- 6.4 CNN -- 6.5 ResNet -- 6.6 RCNN -- 6.7 Comparing of Models -- 6.8 New Model -- 6.9 Summarization -- 7 Conclusion -- References -- Combined Feature Representation for Emotion Classification from Russian Speech -- Abstract -- 1 Introduction and Related Work -- 2 Proposed Method -- 2.1 Feature Extraction -- 2.2 Feature Representation -- 2.3 Classification -- 3 Experimental Settings and Results -- 3.1 RUSLANA Database -- 3.2 Experimental Results -- 4 Conclusion -- Acknowledgments -- References -- Information Extraction -- Active Learning with Adaptive Density Weighted Sampling for Information Extraction from Scientific Papers -- 1 Introduction -- 2 Related Work -- 3 Sampling Strategies for Active Learning -- 4 Task-Independent Features and Classification Pipeline -- 5 Annotation Tool -- 6 Experiments -- 6.1 Data -- 6.2 Evaluation Without Active Learning -- 6.3 Evaluation with Active Learning -- 6.4 Corpus Improvement with Active Learning -- 7 Conclusion -- References -- Application of a Hybrid Bi-LSTM-CRF Model to the Task of Russian Named Entity Recognition -- 1 Introduction -- 2 Neuronal NER Models -- 2.1 Long Short-Term Memory Recurrent Neural Networks -- 2.2 Bi-LSTM -- 2.3 CRF Model for NER Task -- 2.4 Combined Bi-LSTM and CRF Model -- 2.5 Neuro NER Extensions -- 3 Experiments -- 3.1 Datasets -- 3.2 External Word Embedding -- 3.3 Results -- 4 Discussion -- 5 Conclusions -- References -- Web-Scale Data Processing -- Employing Wikipedia Data for Coreference Resolution in Russian -- Abstract -- 1 Introduction -- 2 Related Work on Topic. 3 Using Semantic Features from Wikipedia Data to Improve Results of Coreference Resolution -- 3.1 Text Preprocessing and Feature Extraction -- 3.2 Adding Wikipedia Data -- 4 Results -- 4.1 Discussion -- 4.2 Future Work -- References -- Building Wordnet for Russian Language from Ru.Wiktionary -- 1 Introduction -- 2 Related Work -- 3 Data -- 4 Algorithm Description -- 4.1 Synonym Relations Extraction -- 4.2 Hierarchical Links Extraction -- 4.3 Links Cleaning -- 5 Results -- 6 Conclusion and Future Work -- References -- Corpus of Syntactic Co-Occurrences: A Delayed Promise -- Abstract -- 1 Online Resources on Word Combinations in Russian -- 2 Method and Used Corpora -- 3 CoSyCo Database and Site -- 4 Evaluation -- 5 Conclusion -- References -- Computation Morphology and Word Embeddings -- A Close Look at Russian Morphological Parsers: Which One Is the Best? --

Abstract -- 1 Introduction -- 2 Previous Work -- 3 Russian Morphological Parsers -- 3.1 Mystem -- 3.2 pymorphy2 -- 3.3 TreeTagger -- 3.4 FreeLing -- 4 Methodology -- 4.1 Corpora -- 4.2 POS Tagsets and Verb Lemmas -- 4.3 Evaluation Measures -- 5 Results and Discussion -- 5.1 Lemmatization -- 5.2 POS Tagging -- 6 Conclusion -- Acknowledgments -- References -- Morpheme Level Word Embedding -- 1 Introduction -- 2 Related Work -- 3 Vocabularies -- 4 Algorithm for Segmentation a Word to Morphemes -- 4.1 Learning -- 4.2 Segmentation -- 5 Morpheme Embedding -- 6 Word Embedding Correction -- 7 Experiments -- 8 Conclusions and Future Work -- References -- Comparison of Vector Space Representations of Documents for the Task of Information Retrieval of Massive Open Online Courses -- 1 Introduction -- 2 Related Work -- 3 Approach to Comparison of Vector Representations -- 3.1 Corpus -- 3.2 Preprocessing -- 3.3 Vector Space Models -- 3.4 Processing Query -- 3.5 Human Judgment -- 3.6 Evaluation Metrics. 4 Results and Discussion -- 5 Conclusion -- References -- Machine Learning -- Interpretable Probabilistic Embeddings: Bridging the Gap Between Topic Models and Neural Networks -- 1 Introduction -- 2 Related Work -- 3 Probabilistic Word Embeddings -- 4 Additive Regularization and Embeddings for Multiple Modalities -- 5 Experiments -- 6 Conclusions -- References -- Multi-objective Topic Modeling for Exploratory Search in Tech News -- 1 Introduction -- 2 Probabilistic Topic Modeling and Additive Regularization -- 3 Topic-Based Exploratory Search -- 4 Experiments with Topic-Based Search -- 5 Model Parameters Optimization -- 6 Conclusions -- References -- A Deep Forest for Transductive Transfer Learning by Using a Consensus Measure -- 1 Introduction -- 2 Deep Forest -- 3 Consensus Measures and Training the TLDF -- 3.1 Weighted Average of Class Probabilities -- 3.2 The Shannon Entropy as a Consensus Measure -- 4 Convex Measure of the Transfer Learning Consistence -- 5 An Algorithm for the TLDF Training -- 6 Numerical Experiments -- 7 Conclusion -- References -- Russian Paraphrase Detection Shared Task -- ParaPhraser: Russian Paraphrase Corpus and Shared Task -- Abstract -- 1 Introduction -- 2 Background -- 2.1 Paraphrase Extraction and Recognition -- 2.2 Paraphrase Corpora -- 3 The ParaPhraser Project -- 3.1 The Construction Process -- 3.2 Crowdsourcing -- 3.3 Evaluation -- 4 Shared Task -- 4.1 The Task -- 4.2 Baselines -- 4.3 Results -- 4.4 Experiments -- 5 Conclusion -- References -- Effect of Semantic Parsing Depth on the Identification of Paraphrases in Russian Texts -- Abstract -- 1 Introduction -- 2 Parser -- 2.1 The SemSin System -- 2.2 Syntactic Parsing and Semantic Analysis Using SemSin -- 3 Text Analysis -- 3.1 Lemmatization -- 3.2 Semantics. Accounting for Classes -- 3.3 Semantics. Synonymy and Additional Classes -- 3.4 Dependency Tree. 4 Results -- 5 Conclusion -- References -- RuThes Thesaurus in Detecting Russian Paraphrases -- 1 Introduction -- 2 Related Work -- 3 RuThes Thesaurus -- 4 Using RuThes Synonyms in News Article Clustering -- 5 RuThes in Russian Paraphrasing Task -- 5.1 Russian Paraphrasing Task -- 5.2 Evaluating Thesaurus-Based Features in Paraphrase Detection -- 5.3 Finding the Best Thesaurus Feature -- 5.4 Combining Thesaurus Features with Other Features -- 6 Conclusion -- References -- Knowledge-lean Paraphrase Identification Using Character-Based Features -- Abstract -- 1 Introduction -- 2 Related Work -- 3 Paraphrase Corpora -- 3.1 The Microsoft Paraphrase Corpus -- 3.2 The Plagiarism Detection Corpus -- 3.3 The Twitter Paraphrase Corpus -- 3.4 A Turkish Paraphrase Corpus -- 3.5 A Russian Paraphrase Corpus -- 4 Knowledge-Learn Paraphrase Identification --

4.1 Representing Paraphrase Pairs -- 4.2 Classifier Training -- 4.3 Feature Scaling -- 4.4 Experiments -- 5 Combination of Word- and Character-Based Features -- 6 The Russian Paraphrase Task -- 6.1 Three Class Versus Binary Classification -- 6.2 Results -- 7 Discussion and Conclusions -- Acknowledgements -- References -- Paraphrase Detection Using Machine Translation and Textual Similarity Algorithms -- 1 Introduction -- 1.1 Motivation -- 1.2 Objective -- 1.3 Task Description -- 2 Related Work -- 3 Data Set -- 4 Baseline -- 4.1 Algorithm -- 4.2 Results -- 5 Algorithm -- 5.1 Brief Explanation -- 5.2 Detailed Description -- 5.3 Feature Vector Structure for Each One of the Three Translations -- 6 Comparison of Toolkits on First Task (3-Way Classification) -- 6.1 Results -- 6.2 Confusion Matrix -- 7 Ablation Test and Its Analysis on Second Task (2-Way Classification) -- 7.1 Results -- 7.2 Confusion Matrix -- 7.3 Identifying Best SEMILAR Toolkit Score.  
8 Comparison of Translation Engines for Second Task (2-Way Classification).

---

Sommario/riassunto

This book constitutes the refereed proceedings of the 6th Conference on Artificial Intelligence and Natural Language, AINL 2017, held in St. Petersburg, Russia, in September 2017. The 13 revised full papers, 4 revised short papers papers were carefully reviewed and selected from 35 submissions. The papers are organized in topical sections on social interaction analysis, speech processing, information extraction, Web-scale data processing, computation morphology and word embedding, machine learning. The volume also contains 6 papers participating in the Russian paraphrase detection shared task.

---