Dmitry Frolov

Curriculum Vitae

Contact Information

Home address: Moscow, Russia

Telephone number: +7-926-674-9000

E-address: dmitsf@gmail.com

Education

Ph.D. in Computer Science, 2014-2019, Department of Data Analysis and Artificial Intelligence, Faculty of Computer Science, HSE University (Moscow, Russia). Research Supervisor: Prof. Boris Mirkin. Dissertation: Aggregate representation of texts for information retrieval.

M.S. in Software Engineering, 2012-2014, Department of Software Engineering, Faculty of Computer Science, HSE University (Moscow, Russia). Research Supervisor: Prof. Boris Mirkin.

B.S. in Applied Mathematics and Computer Science, 2008-2012, Department of Mathematics. Petrozavodsk State University (Petrozavodsk, Russia). Graduated with distinction and honors.

Skills

Programming Languages: Python (2/3); C; C++; JavaScript; Unix Shell

Data Science & ML Tools: NLTK; PyMorphy; Scikit-Learn; Gensim; Pandas; MLLib; Statsmodels; Keras; Tensorflow

Databases & Storages: MongoDB; Aerospike; PostgreSQL; Redis

Other Technologies: Git; Docker; AWS, etc.

Work Experience

Oct. 2015 – present: Lead Developer (Data Science and Machine Learning), Chief Technical Officer (from Jan. 2018), Natimatica Ltd., Moscow, Russia

Selected Projects:

- Developed and implemented advanced predictive advertisement targeting models (~2.5 times CTR improved; ~2.1 times revenue increased. Tools used: Python; Scipy, Numpy, Statsmodels, Scikit-Learn; Docker, AWS.
- Designed and implemented Data Management Platform (data of more than 90M active users stored). Tools used: Apache Kafka, Apache Spark, Aerospike, MongoDB; Python.
- Developed and implemented an online user interest classifier (using IAB taxonomy). Tools used: Python; NLTK, Gensim, Scikit-Learn.
- Developed and implemented user audience extending tool (based on my PhD project results). Tools used: Python, C.
- Designed “Ad-on-Image” technical scheme, managed image recognition system development. Tools used: Python; Keras, Tensorflow.
- Designed and implemented a recommender system and basic advertisement targeting modules. Tools used: Python, MLLib, FMLib, Numpy, Scipy, Scikit-Learn.

- Designed and implemented website content parsers and natural language processing modules. Tools used: Python, Scrapy, NLTK, PyMorphy, Gensim, Yandex MyStem.
- Hired Data Science and Engineering teams.

Sep. 2012 - Oct. 2015: Software Developer (Data platform), Lead Developer (from Jan. 2015), SlickJump Ltd., Moscow, Russia

Selected Projects:

- Designed and implemented a multicriterial advertising platform.
- Implemented advertising billing system.
- Implemented advertising backend, designed client-server interaction.
- Designed and implemented natural language processing modules.

Feb. 2010 - May 2012: Software Engineer, Laboratory of IT Systems, Petrozavodsk University. Petrozavodsk, Russia

Research Experience

Feb. 2016 – present: Research Assistant, International Laboratory of Decision Choice and Analysis. HSE University (Moscow, Russia).

- Work with Prof. Boris Mirkin on research project "Finding an appropriate generalization for a fuzzy thematic set in taxonomy."
- Member of funded "Concept" Research Group under Prof. Boris Mirkin supervision.

Nov. 2016 - Dec. 2016: Research Assistant (Visitor), Department of Computer Science and Information Systems, Birkbeck University of London. London, UK.

- Worked on research project "Automating Annotation of Research Papers Using the ACM Classification of Computing Systems."

Teaching Experience

Jan. 2015 - Feb. 2018: Lecturer, Mentor. HSE University (Moscow, Russia), Department of Data Analysis and Artificial Intelligence. Courses: "Data Analysis," "Unstructured Data Analysis."

Judgement and Reviewing Experience

SN Computer Science Journal. Reviewer in Machine Learning section (Jan. 2020 – present).

The 3rd International Conference on Computer Science and Application Engineering (CSAE). China, Sanya. Reviewer (2019, 2020).

DDA 2018 Conference. Russia, Moscow, 27-28 Dec. 2018, Member of Organizing Committee.

International NRU HSE Olympiad for Graduates, Computer Science. 2015, 2016, 2018. Member of Organizing Committee.

All-Russian Mathematical Olympiad. 2011. Member of Commission.

Professional Committees

Russian Association for Artificial Intelligence (RAAI). Member.

Publications

Frolov, D., Nascimento, S., Fenner, T. and Mirkin, B., 2019. Parsimonious Generalization of Fuzzy Thematic Sets in Taxonomies Applied to the Analysis of Tendencies of Research in Data Science. *Information Sciences*. DOI: http://doi.org/10.1016/j.ins.2019.09.082

D. Frolov, S. Nascimento, T. Fenner, B. Mirkin, 2019. Using Taxonomy Tree to Generalize a Fuzzy Thematic Cluster. *FUZZ IEEE Proc*., IEEE. DOI: http://doi.org/10.1109/FUZZ-IEEE.2019.8859015

Frolov D., Mirkin B., Nascimento S., Fenner T., 2019. Method for Generalization of Fuzzy Sets. *ICAISC Proceedings*, *LNCS*, Springer. P. 273-286. DOI: http://doi.org/10.1007/978-3-030-20912-4

D. Frolov, S. Nascimento, T. Fenner, Z. Taran, B. Mirkin, 2019. Computational Generalization in Taxonomies Applied to: (1) Analyze Tendencies of Research and (2) Extend User Audiences. *IDEAL 2019 Proceedings*, Springer. DOI: https://doi.org/10.1007/978-3-030-33617-2_1

Dmitry Frolov, Zina Taran, Boris Mirkin, 2019. A Method for Audience Extending in Programmatic Advertising by Using Parsimonious Generalization of User Segments. *IHIET-2019 Proceedings*, AICS, Springer. P. 837-841. DOI: http://doi.org/10.1007/978-3-030-25629-6

D. Frolov, B. Mirkin, S. Nascimento, T. Fenner, 2019. Globally Optimal Parsimoniously Lifting a Fuzzy Query Set Over a Taxonomy Tree, *WCGO-2019 Proceedings. LNCS*, Springer. P. 779-789. DOI: http://doi.org/10.1007/978-3-030-21803-4

D. Frolov, B. Mirkin, S. Nascimento, T. Fenner, 2019. Using Domain Taxonomy to Model Generalization of Thematic Fuzzy Clusters. *IARIA Content 2019 Proceedings*.

D. Frolov, B. Mirkin, S. Nascimento, T. Fenner, 2018. Finding an appropriate generalization for a fuzzy thematic set in taxonomy. Working paper WP7/2018/04, Moscow, Higher School of Economics Publ. House.

Dmitry Frolov, 2016. Using Annotated Suffix Trees for Fuzzy Full Text Search. *RuSSIR 2016 Proceedings*. CCIS, Springer.

Frolov D.S., 2015. Annotated suffix tree as a way of text representation for information retrieval in text collections. *Business Informatics*. No. 4/2015. P. 63-70. DOI: https://doi.org/10.17323/1998-0663.2015.4.63.70

B. Mirkin, D. Frolov, A. Vlasov, S. Nascimento, and T. Fenner. 2020. A Hybrid Approach to Interpretable Analysis of Research Paper Collections. In *The 10th International Conference on Web Intelligence, Mining and Semantics (WIMS 2020)*, June 30-July 3, 2020, Biarritz, France. ACM. https://doi.org/10.1145/3405962.3405976

Mirkin B., Frolov D., Vlasov A., Nascimento S., Fenner T. (2020) A Hybrid Approach to the Analysis of a Collection of Research Papers. In *Intelligent Data Engineering and Automated Learning – IDEAL 2020*. Lecture Notes in Computer Science, vol 12490. Springer, Cham. https://doi.org/10.1007/978-3-030-62365-4_40

Conferences

Speaker: 21st International Conference on Intelligent Data Engineering and Automated Learning - IDEAL 20204th-6th November 2020, Guimarães, Portugal *A Hybrid Approach to the Analysis of a Collection of Research Papers.*

Speaker: 10th International Conference on Web Intelligence, Mining and Semantics (WIMS-2020), France, Biarritz. *A Hybrid Approach to Interpretable Analysis of Research Paper Collections.*

Speaker: 2nd International Conference on Artificial Intelligence towards Industry 4.0 (ICAII4.0-2019), Turkey, Iskenderun. *Efficient Audience Extending in Targeted Advertising by Using Generalization of User Segments.*

Speaker: 6th World Congress on Global Optimization (WCGO-2019), France, Metz. *Globally Optimal Parsimoniously Lifting a Fuzzy Query Set Over a Taxonomy Tree.*

Poster: The 18th International Conference on Artificial Intelligence and Soft Computing (ICAISC-2019), Poland, Zakopane. *Method for Generalization of Fuzzy Sets*.

Speaker: All-Moscow scientific seminar “Mathematical methods of decision analysis in economics, business and politics,” Moscow, Russia, 2018. *Annotation of a Document Collection by Finding Thematic Fuzzy Clusters and Parsimoniously Lifting Them in a Domain Taxonomy*.

Poster: 3rd Kolmogorov seminar on computer linguistics and language sciences, Moscow, Russia, 2018. *Annotation of a Document Collection by Finding Thematic Fuzzy Clusters and Parsimoniously Lifting Them in a Domain Taxonomy*.

Poster: Russian School of Information Retrieval (RuSSIR) - 2016, Saratov, Russia. *Using Annotated Suffix Trees for Fuzzy Full Text Search.*

Poster: Summer School of Faculty of Computer Science, HSE University, Moscow, Russia, 2016. *Using Annotated Suffix Trees for Fuzzy Full Text Search.*

Poster: Russian School of Information Retrieval (RuSSIR) - 2015. Saint-Petersburg, Russia. *Aggregate Text Representation for Information Retrieval in Collections of Text Documents.*

Network activity

Github: https://github.com/dmitsf

Google Scholar: https://scholar.google.ru/citations?user=v8qV5rUAAAAJ

StackOverflow: https://stackoverflow.com/users/3824945/dmitry