

Siarhei Bykau

Purdue University
Department of Computer Science
305 N. University Street West Lafayette
IN 47907, USA

+1 (765) 586-9524
sbykau@purdue.edu
<http://www.cs.purdue.edu/homes/sbykau/>

AREAS OF EXPERTISE/SKILLS

Expertise: Machine Learning, Probabilistic Graphical Models, Statistics, Relational Databases, Recommendation Systems, Algorithms, Data Mining, Web Development, Visualization.

Skills: Java, Python, SQL, MySQL, Oracle, GitHub, Amazon Web Services, pandas, matplotlib, graphviz, ipython notebook, xgboost, random forest, SVM, J2EE, Tomcat, Struts, Hibernate, Google Web Toolkit

DATA SCIENCE PROJECTS

Predict Next Destination for Airbnb Users

Description: The project predicts in which country a new Airbnb user will make his or her first booking based on their demographics, web session records, and summary statistics. The models are trained and validated on AWS. The final model scored top 10% at Kaggle.

Tools: *python, amazon web services, pandas, matplotlib, xgboost, random forest, github, ipython notebook*

Controversy Detection in Wikipedia

Description: The project finds disputes in the Wikipedia history dump (7TB) by clustering n-grams of words and links. The results are stored in MySQL and visualized as a web application.

Tools: *java, sql, clustering, google web toolkit, crowdsourcing*

Mining Corporate History with US Trademark datasets

Description: The project extracts the corporate history of all US companies based on the USPTO trademark dataset. Using a web application, the found graphs are visualized and can be queried.

Tools: *data mining, java, sql, dynamic programming, visualization, graphviz, algorithms, google web toolkit*

Cold Start Movie Recommendations

Description: The project implements a number of machine learning algorithms to recommend movies based on their features (e.g. actors, genre). The models are trained and evaluated on the Movie Lens dataset.

Tools: *collaborative filtering, apache mahout, java, visualization*

Web Data Fusion

Description: The project implements statistical algorithms to fuse contradictory information from Web sites (e.g. conflicting flight departure times). It models the problem as a graphical model and does the inference with expectation-maximization algorithm.

Tools: *machine learning, statistics, expectation maximization, graphical models, python, ipython notebook*

Contextual Linkage

Description: The project links contextually relevant links on Wikipedia (7TB). It uses decision trees and correlation clustering to track relevant entities across data snapshots. The results are validated with Amazon Mechanical Turk.

Tools: *decision trees, correlation clustering, python, matplotlib*

EDUCATION

Postdoctoral Associate in Computer Science, Purdue University Jan 2014 - Now

West Lafayette, IN, USA

PhD in Computer Science, University of Trento Nov 2008 - Apr 2013

Trento, Italy

M.Sc. in E-Government Technologies, University of Trento Oct 2007 - Jul 2008

Trento, Italy

B.Sc., Radio Physics and Electronics Faculty, Belarusian State University 2002 - 2007

Minsk, Belarus

PROFESSIONAL EXPERIENCE

Big Data Developer, Trento Rise

May 2013 - Dec 2013

Trento, Italy

Designed a topical crawler based on Apache Nutch.

Intern, AT&T Labs Inc., Shannon Lab

May 2011 - Nov 2011

NJ, USA

Designed, developed the framework of controversy detection in Wikipedia.

Java Developer, Artezio LLC

May 2007 - Oct 2007

Minsk, Belarus

Developed a system for online medical image management for Perceptive Informatics.

Java Developer, EPAM Systems

Jul 2005 - May 2007

Minsk, Belarus

Developed ranking algorithms and authentication modules the search engine of the SAP NetWeaver platform.

PUBLICATIONS

- **Bykau, S.**, Lee, J., Srivastava D., Velegrakis, Y. Contextual Entity Linkage. (VLDB) (submitted) , 2016
- Pradhan, R., **Bykau, S.**, Prabhakar, S. A Human-Machine approach to Pay-as-you-go Truth Discovery. (KDD) (submitted) , 2016
- **Bykau, S.**, Korn, F., Srivastava, D., Velegrakis, Y. Controversy Detection in Wikipedia. International Conference on Data Engineering (ICDE), 2015
- **Bykau, S.**, Koutrika, G., Velegrakis, Y. On Dealing with the Permanent Coping with the Persistent Cold-start Problem. Personalized Access, Profile Management, and Context Awareness in Databases (PersDB), 2013
- **Bykau, S.**, Rizzolo, F., Velegrakis, Y. A Query Answering System for Data with Evolution Relationships. SIGMOD Conference (SIGMOD), 2013
- **Bykau, S.**, Mylopoulos, J., Rizzolo, F., & Velegrakis, Y. On Modeling and Querying Concept Evolution. Journal on Data Semantics (JoDS), 2012
- **Bykau, S.**, Mylopoulos, J., Rizzolo, F., & Velegrakis, Y. Supporting Queries Spanning Across Phases of Evolving Artifacts using Steiner Forests. International Conference on Information and Knowledge Management (CIKM), 2011
- Katifori, A., Nikolaou, C., Platakis, M., Ioannidis, Y., Tympas, A., Koubarakis, M., Sarris, N., Tountopoulos, V., Tzoannos, E., **Bykau, S.**, Kiyavitskaya, N., Tsinaraki, C., Velegrakis, Y. The Papyrus digital library: discovering history in the news. International Conference on Theory and Practice of Digital Libraries (TPDL), 2011
- **Bykau, S.**, Kiyavitskaya N., Tsinaraki C., Velegrakis Y. Bridging the gap between heterogeneous and semantically diverse content of different disciplines. International Workshop on Flexible Database and Information System Technology (FlexDBIST), 2010.
- Rizzolo F., Velegrakis Y., Mylopoulos J., and **Bykau S.** Modeling concept evolution: a historical perspective. International Conference on Conceptual Modeling (ER), 2009.
- Presa, A., Velegrakis Y., Rizzolo R., and **Bykau S.** Modelling Associations through Intensional Attributes. International Conference on Conceptual Modeling (ER), 2009.

TEACHING EXPERIENCE

Instructor for Introduction to Relational Database Systems
Computer Science Department, Purdue University, USA

Spring'15

Teaching Assistant for Information Managements and Integration Fall08, Fall09, Fall10, Fall11
Department of Information and Telecommunications, University of Trento, Italy

Teaching Assistant for Advance Information Systems
Department of Information and Telecommunications, University of Trento, Italy

Spring08, Spring09

HONORS, AWARDS, ACHIEVEMENTS AND GRANTS

- Airbnb New User Bookings, top 10%, Feb 2016
- Kaggle Acquire Valued Shoppers Challenge, top 10%, July 2014
- PhD Scholarship, University of Trento, Nov 2008 – Apr 2013
- The highest final mark (107 out of 110) at Master program, Jul 2008
- Intern Scholarship, University of Trento, May 2008 – Jul 2008
- Winner (7th place) of BSU Olympiad in physics, May 2002
- Winner (3d place) of Olympiad in physics (Vitebsk region), Jan 2002
- Honorable mention at Republic Olympiad in physics, Mar 2001
- Winner (1st place) of Olympiad in physics (Vitebsk region), Jan 2001

MISCELLANEOUS

GitHub: <https://github.com/bykau>

Kaggle: <https://www.kaggle.com/sbykau>

Topcoder: <https://www.topcoder.com/members/cyrrix/>

Google Scholar: <https://scholar.google.com/citations?user=dF07rwsAAAAJ>

Languages: English (fluent), Italian (intermediate), Russian (native)

Chess: ELO Rating 2000