

# Evgeny Karataev

---

<b>CONTACT INFORMATION</b>	School of Information Sciences University of Pittsburgh	phone: 412 759 7107 email: Karataev.Evgeny@gmail.com web: evgenykarataev.com
<b>RESEARCH INTERESTS</b>	I am interested in Data Management, Data Integration, and Distributed Systems. My recent research has focused on developing large-scale data integration infrastructure, spatiotemporal approximate join and keyword search over virtually integrated heterogeneous distributed datasets.  I like to do both research and software development, but even more I like to combine them together by working on innovative projects which require cutting edge technologies and result in products that go beyond academic environment.	
<b>EDUCATION</b>	<b>University of Pittsburgh</b> , Pittsburgh, PA, USA PhD, Information Sciences and Technology Advisor: Vladimir Zadorozhny Committee: Marek Druzdzel, Konstantinos Pelechrinis, Nicholas Nystrom Dissertation: <i>Advanced Distributed Data Integration Infrastructure and Research Data Management Portal</i> GPA: 3.98  <b>Tomsk State University of Control Systems and Radioelectronics</b> , Tomsk, Russia Dipl.-Ing. (Cum Laude) major in Computer Aided Design at Computing System Department GPA: 4.94/5.	December 2016        2011
<b>PROFESSIONAL EXPERIENCE</b>	<b>Siemens, Corporate Core Division</b> Intern - Evaluation of Intelligent System Building Blocks Worked on developing and evaluating intelligent algorithms, particularly worked on genetic algorithms for train scheduling task: <ul style="list-style-type: none"><li>• Refactored, reorganized and “mavenized” java code base to improve CI (added ~100 test)</li><li>• Extended domain model, so more real life use cases can be addressed</li><li>• Optimized genetic algorithms project by minimizing IO operations, improving data structures and code practices, as the result performance grew by 18 times</li><li>• Identified and fixed over 10 critical bugs</li><li>• Developed two new mutation operators to explore more options in solution space</li><li>• Created benchmarking project and abstractions to allow easy way to add new benchmarking scenarios. Developed GA profiler to capture execution time of various GA computation steps and generated comprehensive json output with all important information.</li></ul> Used technologies and tools: Java, Python, GA Watchmaker Framework, Git, Phabricator, Jenkins, Shell, Ubuntu, NVIDIA DIGITS	May – July 2016
	<b>University of Pittsburgh</b> Graduate Student Assistant Working as Teaching Assistant for graduate level "Database Management Systems" course.	Fall 2015
	<b>University of Pittsburgh</b> Instructor/Teaching Fellow  Taught a graduate level Advanced Topic in Database Management class. Topics that are covered include: Data Integration (OLAP and Data Warehousing, Virtual Data Integration), Distributed and Parallel Databases (including distributed transactions and query execution), NoSQL databases, NewSQL databases (Main Memory Databases), Cluster Computing (Hadoop and other animals, Spark)	January – April 2015

	<b>Siemens Corporation, Corporate Research and Technology</b>	May –
	Software Engineer – Intern	December 2014
	<ul style="list-style-type: none"> <li>• Developed web application for software code and architecture analysis.</li> <li>• Reduced memory and time complexity to extract, transform and integrate artifact models by orders of magnitude by developing mutli-threaded stream-based algorithms.</li> <li>• Implemented Clang AST, Understand and BTrace models extraction, transformation and integration.</li> </ul>	
	Used technologies and tools: Java, Python, ExtJs, Git, Phabricator, Jenkins, Shell, Understand, Clang, MinGW, Eclipse EMF, Eclipse CDO, Neo4j, H2, Hibernate, MongoDB	
	<b>University of Pittsburgh</b>	Summer 2013
	Instructor/Teaching Fellow	
	Taught graduate level Database Management Systems class	
	<b>University of Pittsburgh</b>	2013-2014
	Graduate Research Assistant	
	Worked on NSF funded research project as a researcher, lead developer and project manager:	
	<ul style="list-style-type: none"> <li>• Developed global-scale advanced infrastructure for data integration, preservation and analysis.</li> <li>• Administered development (with continuous integration and testing) and production servers.</li> <li>• Supervised over 60 master students</li> </ul>	
	<b>Siemens Corporation, Corporate Research and Technology</b>	Summer 2012
	Software Engineer – Intern	
	Worked on Power Distribution management project - Demand Response Management System (DRMS) and was primarily in charge of the GUI that included asset management and user task support, visualization of various aspects of system status and control, and support for visualizing analytics and prediction from the available load data history in the future releases.	
	One of the outcomes of my work was an improved tree view that could scale to millions of nodes. The work led to invention disclosure and US patent titled "Method and apparatus for browsing large data network topology trees"	
	Used technologies and tools: Java, ActionScript, Flex, Mercurial, Jira, Bamboo	
	<b>University of Pittsburgh</b>	2011 - 2013
	Graduate Student Assistant	
	Worked as Teaching Assistant for "Database Management Systems", "Introduction to Information Technology" and "Introduction To Information, Systems and Society" courses	
	<b>Tomsk State University of Control Systems and Radioelectronics (TUSUR)</b>	2009 - 2011
	Laboratory of Intelligent Computer Systems	
	Software developer/Research Assistant	
	Worked as a researcher and software developer on Indesys System - intelligent CAD system for automated synthesis of microwave devices; contributed to preparation of Federal Target Program grant “Designing of methods, algorithms and intelligent software for synthesis microelectronic SHF-devices by the use of precise models of integral elements”	
<b>PATENT</b>	Method and apparatus for browsing large data network topology trees United States Patent WO2014036073 A2 Inventors: Gilberto Matos, Evgeny Karataev	Issued August 28, 2013
<b>PUBLICATIONS</b>	<b>Evgeny Karataev</b> , Vladimir Zadorozhny. <i>Temporal Aggregate Join</i> (submitted to ICDE 2017)	
	Vladimir Zadorozhny, Manning Patrick, Matthew Drwenski, and <b>Evgeny Karataev</b> . <i>Towards a Social Weather Service: Linking Social Sciences and Information Sciences in Large-Scale Inequality Analysis</i> . Journal of World-Historical Information 2, no. 1, 2016	
	<b>Evgeny Karataev</b> , Vladimir Zadorozhny. <i>Adaptive Social Learning Based On Crowdsourcing</i> . IEEE Transactions on Learning Technologies, Jan. 2016 (DOI: 10.1109/TLT.2016.2515097)	

**Evgeny Karataev**, Vladimir Zadorozhny. *Col\*Fusion: a Global-Scale Information Integration Infrastructure based on Crowdsourcing*. Journal Of World-Historical Information, Vol. 2-3 (2014-2015), No. 1, 2015

**Evgeny Karataev**, Vladimir Zadorozhny. *Col\*Fusion: Scalable Information Integration Infrastructure based on Crowdsourcing for Data-Driven Collaboration in History*. 39th Annual Meeting of the Social Science History Association, Canada, 2014

**Evgeny Karataev**, Vladimir Zadorozhny. *Col\*Fusion: not just yet another data repository*. *iConference*, 2014

Andrei Salnikov, Igor Dobush, Alesya Stepacheva, **Evgeny Karataev**, Alexey Abramov. *Program for automation of measurements and statistical analysis of elements parameters of microwave monolithic integrated circuits based on the Indesys-MS software*. All-Russian scientific-practical conference "Contemporary problems in radio electronics", SFU, Krasnoyarsk, 2012

Andrei Salnikov, **Evgeny Karataev**, Igor Dobush. *Software Programs for Storage and Statistical Analysis of MMIC Measurement Data*. Microwave and Telecommunication Technology (CriMiCo), 21th International Crimean Conference, 2011 (In Russian)

**Evgeny Karataev**, Andrei Salnikov. *Software module as a part of Indesys-MS for storing and statistical analysis of microwave measurements*. XVI International Scientific and Practical Conference of Students and Young Scientists "Modern Technique and Technology 2011", Tomsk Polytechnic University, 2011 (In Russian)

**Evgeny Karataev**, Michail Peskov, Sergey Dorofeev. *Platform Indesys Framework for the creation of specialized programs in the design of microwave devices*. XVI International Scientific and Practical Conference of Students and Young Scientists "Modern Technique and Technology" / Collected Works in 3 volumes. V. 3. – Tomsk: Publishing House of Tomsk Polytechnic University, 2010 (In Russian)

**Evgeny Karataev**, Michail Peskov, Sergey Dorofeev. *Data visualization in system for computer-aided synthesis Indesys*. Festival of Microsoft in the TPU. Session "Intelligent Systems and Technologies» VII All-Russian scientific-practical conference of students, graduates and young scientists "Microsoft Technologies in theory and practice of programming", 2010 (In Russian)

**Evgeny Karataev**, Michail Peskov, Sergey Dorofeev. *Versatile data visualization module*. Proceedings of the All-Russian Scientific Conference of Students and Young Scientists "Scientific session TUSUR-2010", 2010 (In Russian)

**Evgeny Karataev**, Sergey Dorofeev. *Module for graphical representation of numerical data in the system Indesys*. XLVIII International Student Scientific Conference "Student and Scientific-Technical Progress" 2010 (In Russian)

## TALKS

**Advanced distributed data integration infrastructure and research data management portal** 2016

Siemens, Princeton, NJ

**What's new in Col\*Fusion** 2014

World Historical Gazetteer workshop organized by CHIA, Pittsburgh, PA

**Col\*Fusion: Global Scale Data Integration Infrastructure** 2013

CHIA Workshop, Pittsburgh, PA

**SALT: Self-Adaptive Learning through Teaching** 2013

PAWS Lab, University of Pittsburgh

## PROJECTS

**Fuse Join** 2015 - on

Developing approximate and exact techniques to fuse datasets based on string, spatial and temporal attributes. Developed novel name-hierarchy based spatial join and temporal aggregate join methods.

Used technologies and tools: Java, Matlab, Pittsburgh Supercomputer Center

	<p><b>MultiDBs</b> Query, keyword search and analysis over heterogeneous databases via unified interface.</p> <p>Used technologies and tools: Java, HTML, JavaScript, CSS, Knockout.js, REST API, Jersey, Jetty, Neo4j, PrestoDB, MySQL, Maven, Docker, AWS, IPython Notebooks (Jupyter Hub), Git, Github, Ubuntu Server</p>	Spring 2015
	<p><b>Col*Fusion: Collaborative Data Fusion</b> Collective data integration, preservation and analysis based on crowdsourcing Techniques.</p> <p>Used technologies and tools: Java, PHP, HTML, JavaScript, CSS, Pligg, Smarty Templates, Knockout.js, jQuery, RESP API, Jersey, Apache2 HTTP Server, MySQL, PostgreSQL, MS SQL, Neo4j, Docker, Vagrant, Pittsburgh Supercomputer Center, Maven, Jenkins, Redmine, Git, Windows and Ubuntu Servers</p>	2013 - on
	<p><b>SALT: Self-Adaptive Learning through Teaching</b> Educational social network to share knowledge and learn from others with recommendation engine.</p> <p>Used technologies and tools: C#, ASP.NET MVC, MS SQL Server, IIS, SignalR, Lucene.Net, d3.js, SVN</p>	2011 - 2012
	<p><b>Indesys MS - Storage and Statistics Module</b> A module for storing measurements of parameters of elements of microwave monolithic integrated circuits in a database and for statistical analysis and advanced visualization thereof.</p> <p>Used technologies and tools: C#, GDI+, MySQL, SVN</p>	2010 - 2011
	<p><b>Indesys - Visualization module</b> User controls for displaying graphs of functions in different coordinate systems (Rectangular, Smith chart, Polar chart, Table).</p> <p>Used technologies and tools: C#, GDI+, SVN</p>	2009 - 2010
	<p><b>3D simulation of the electrostatic energy-analyzer</b> Interactive 3D simulation of the electrostatic energy-analyzer of beam of charged particles.</p> <p>Used technologies and tools: Delphi, OpenGL</p>	2005 - 2006
<b>HONORS AND AWARDS</b>	<p><b>Research Poster Competition</b>, University of Pittsburgh Awarded 1st prize out of 8 projects</p>	2014
	<p><b>“Imagine Cup”</b> held by Microsoft, Siberia Region Stage Awarded 2nd prize out of 9 projects</p>	2010
	<p><b>“Best project”</b>, Tomsk State University of Controls Systems and Radioelectronis Awarded 2nd prize out of 23 projects in the competition between students group-based projects</p>	2010
	<p><b>“Hero of TUSUR 2009”</b>, Tomsk State University of Controls Systems and Radioelectronis Outstanding achievements diploma in studies and science among 3rd year students.</p>	2010
	<p><b>Gold Medal and First Prize Award</b> on International Festival of Scientific Project among high-school students, Rethymno, Greece</p>	2005
	<p><b>Best Project in Physics SPIE Award</b>, Republican science competition, Astana, Kazakhstan</p>	2005

<b>TECHNICAL SKILLS</b>	<p>Java; C#; JavaScript; Python; PHP; HTML; SQL; C; C++; Matlab; Haskell; Prolog; Action Script.</p> <p>RESTful Web Services; ASP.NET MVC; Django; WinForms; WPF; EF; Hibernate; Flex; ExtJs; OpenGL; Polymer Project.</p> <p>Docker; Vagrant.</p> <p>MySQL; MS SQL Server; PostgreSQL; Oracle; Neo4j, MongoDB; Riak.</p> <p>Apache Mahout; Hadoop; Hive; Spark; PrestoDB; Pentaho DI.</p> <p>SVN; Mercurial; Git; Redmine; JIRA; TFS; Jenkins.</p>	
<b>SERVICES</b>	<p>Webmaster for MDM 2015 conference</p> <p>Sub-reviewer: WISE 2012, MDM 2013</p>	
<b>LANGUAGES</b>	Russian (Native), English (Fluent), Chinese (Beginner), German (Beginner)	
<b>EXTRA CURRICULAR ACTIVITIES</b>	<p>Business Manager of the International Students Organization</p> <p>School of Information, Sciences, University of Pittsburgh</p>	2011 - 2013
<b>MISCELLANEOUS</b>	<p>Short-term language studies diploma, Chinese Language</p> <p>Shanghai International Studies University</p> <p>High School Diploma</p> <p>Regional Specialized Boarding School for Gifted children "Murager"</p> <p>GPA: 4.9/5</p> <p>Diploma of Music Education, Classical Guitar</p> <p>Children's Music School #1</p> <p>Second place on the Republic Music Contest</p>	<p>2010 - 2010</p> <p>2002 - 2006</p> <p>1998 - 2004</p> <p>2003</p>
<b>REFERENCES</b>	Available upon request.	

Last updated on October 22, 2016