

# Paul Suganthan G. C.

Madison, WI - 53706

(608)-960-0666

paul.suganthan@gmail.com

linkedin.com/in/paulgc

---

INTERESTS	Data Management, Data Integration, Data Science, Big Data, Machine Learning, Crowdsourcing.
EDUCATION	<b>University of Wisconsin-Madison</b> August 2012 - Present Ph.D. Computer Sciences Advisor: Prof. AnHai Doan <ul style="list-style-type: none"><li>Dissertation: Toward Effective Workflow Development and Execution for Entity Matching</li></ul> M.S. Computer Sciences <b>GPA: 3.94 / 4.0</b> <b>College of Engineering Guindy, Anna University, India</b> July 2008 - May 2012 B.E. Computer Science and Engineering <b>GPA: 9.77 / 10.0</b> <ul style="list-style-type: none"><li>Thesis: Search Engine Enhancement by Extracting Hidden AJAX Content in Web Applications</li></ul>
WORK EXPERIENCE	<i>Research Assistant</i> , CS. Dept. Aug 2012 - Present <b>University of Wisconsin-Madison</b> <ul style="list-style-type: none"><li>Primary research focuses on developing techniques to scale execution of Entity Matching (EM) workflows containing rules, machine learning (ML), and crowdsourcing operations.</li><li>Scale execution of ML models over the join of two tables using an RDBMS style approach.</li><li>Help domain scientists perform EM by developing a scalable “hands-off” crowdsourced EM solution. System deployed as a service at CloudMatcher.io.</li><li>Monitoring real-time events in Twitter using rules, knowledge base, and ML.</li></ul> <i>Open Source Developer</i> Jan 2016 - Present <ul style="list-style-type: none"><li>Main developer of two Python packages providing tools for scalable string matching (<i>py_stringmatching</i> and <i>py_stringsimjoin</i>). Managed the end-to-end development and release process.</li><li>Supervised graduate student contributors by reviewing code, and guiding them on best practices.</li><li>Packages are currently being used at multiple organizations and in data science classes.</li></ul> <i>Software Engineering Intern</i> , Ads Infrastructure May 2015 - Aug 2015 <b>Google, Mountain View</b> <ul style="list-style-type: none"><li>Building a Natural Language Interface to Databases. Worked on semantic analysis of the natural language query and generation of SQL.</li><li>Developed a failure handling mechanism, which helps in handling ambiguous natural language queries by trying out different interpretations of the query.</li></ul> <i>Software Engineering Intern</i> , Product Classification May 2014 - Aug 2014 <b>Walmart Labs, Mountain View</b> <ul style="list-style-type: none"><li>Worked on automatically generating rules for Product classification and optimizing the execution of such rules. Resulted in a SIGMOD 2015 industrial track paper.</li><li>Developed an interactive tool to help analysts write, refine, and manage regex based classification rules.</li></ul> <i>Teaching Assistant</i> , CS Dept. University of Wisconsin-Madison Jan 2015 - May 2015 <ul style="list-style-type: none"><li>Course TA for the <i>Data Science</i> course.</li></ul>
SELECTED PUBLICATIONS	<ul style="list-style-type: none"><li><i>Smurf: String Similarity Joins Using Random Forest Conditions</i> (Under Submission)</li><li><i>Falcon: Scaling Up Hands-Off Crowdsourced Entity Matching to Build Cloud Services</i>, SIGMOD 2017</li><li><i>CloudMatcher: A Cloud/Crowd Service for Entity Matching</i>, BigDas Workshop, KDD 2017</li><li><i>Magellan: Toward Building Entity Matching Management Systems</i>, VLDB 2016</li><li><i>Why Big Data Industrial Systems Need Rules and What We Can Do About It</i>, SIGMOD 2015</li></ul>
SOFTWARE SKILLS	<ul style="list-style-type: none"><li>Languages &amp; Tools: Python, Java, C++, Cython, HTML, JavaScript, SQL, Git.</li><li>Data Science Tools: Pandas, Scikit-learn, Numpy, Matplotlib, Dask, Anaconda.</li><li>Experience in Map-Reduce framework — Hadoop.</li></ul>
ACADEMIC ACHIEVEMENTS	<ul style="list-style-type: none"><li>David DeWitt Fellowship, UW Madison 2012 – 2013</li><li>Best Outgoing Undergraduate Student, CEG 2008 – 2012</li></ul>