

jtengle@indiana.edu
<http://www.cs.indiana.edu/~jtengle/>

Jeremy Engle

Indiana University
630-254-5533

RESEARCH INTERESTS

My research interests combine database theory and data mining with a focus on algorithms. My thesis is the MoLS framework which was software engineered to allow every aspect of an approximate functional dependency mining algorithm to be customized using plug-ins. This effort included the development of original algorithms and extension of existing algorithms. This practice of combining existing research and tools with innovation has continued with other projects I have been involved with whether it was data integration or chemistry informatics.

EDUCATION

Indiana University, Bloomington, IN, PhD. Computer Science, Anticipated Summer 2010

Indiana University, Bloomington, IN, M.S. Computer Science, May 2005

DePauw University, Greencastle, IN, B.A. Computer Science major, Psychology minor, 2003

PUBLICATIONS

- Jeremy T. Engle, Edward L. Robertson, and Dimitar G. Nikolov, Ordering Depth First Search to Improve AFD Mining, Technical Report 685, Indiana University (under review CIKM 2010)
- Jeremy Engle, Ying Feng, and Rob Goldstone, Inducing Relatedness Graphs for Data Integration, Technical Report 682, Indiana University
- Jeremy T. Engle and Edward L. Robertson, Depth First Algorithms and Inferencing for AFD Mining, 13th International Database Engineering and Applications Symposium (IDEAS) 2009
- Jeremy T. Engle and Edward L. Robertson, Modularizing Data Mining: A Case Study Framework, Technical Report 678, Indiana University
- Jeremy Engle and Edward Robertson, HLS: Tunable Mining of Approximate Functional Dependencies, 25th British National Conference on Databases (BNCOD) 2008
- MIQIS: Modular Integration of Queryable Information Sources. Catharine M. Wyss, George H.L. Fletcher, Fulya Erdinc, and Jeremy T. Engle. Workshop on Information Integration on the Web (IIWeb), at the 30th Int. Conf. on Very Large Data Bases (VLDB), Toronto, Canada, 30 August 2004, pp. 136-140.

TEACHING EXPERIENCE

Course Instructor, Fall 2009, SQL Programming, A290

Developed 8 week course; selecting textbook, developing course curriculum, conduct lectures, write/grade assignments, and write/grade exams.

Co-course Instructor, Fall 2007, Advanced Database Concepts, B461

Conducted lectures for entire semester and took over instructor's responsibilities for final portion of the semester.

Teaching Assistant

TAs grade, hold office hours, run labs, and cover lectures when instructor is absent.

Graduate/CS Major Undergraduates Courses

- Multiple Semesters, Advanced Database Concepts, B561, Supervisor: Cathy Wyss
- Advanced Database Concepts, B461, Supervisor: Dirk Van Gucht
- Introduction to Software Systems (Java), C212, Supervisor: Sriram Mohan

Non-Major Undergraduates Courses

- Multiple Semesters, Introduction to Programming (Java), A201 and A202, Supervisor: Adrian German
- Introduction to Programming (Alice,Python), A201, Supervisor: Chris Haynes
- Informatics Introduction to Programming (Java), I210, Supervisor: Eric Kisling
- Introduction to Computers and Computing (MS Office), A110

SERVICE

- Steering Committee Student VOICE Project
- Student Transportation Advisory Committee
- Assistant Moderator, Graduate and Professional Student Organization (GPSO)
- Computer Science Representative, GPSO
- Student Representative Dean Search Committee, School of Informatics

PROFESSIONAL EXPERIENCE

HealthComPartners, Mt. Prospect, IL, 2003(summer)-2004(November)

Visual Basic development (VB6, API programming), database development (MS Access), Web programming (HTML, VBscript, JavaScript, ASP).

- Developed beta system to detect arrival of encrypted zip files; decrypt, unzip, move contents to specified locations, and update a database with information.
- Created administration tool to update, maintain database, beta version.
- Designed tool to automate FTP process to client's site.
- Designed and built Access database for backend of search website.
- Created front-end search website for client searches, pdf file viewing based on metadata stored in database. Temporary copies of pdf files are automatically removed at end of user session.
- Graphic editing, HTML development.

THESIS

MoLS: A Framework for Customized Approximate Functional Dependencies Mining

Advisor: Ed Robertson (Dept. of Computer Science Indiana University)

Algorithm Development

- Created Lozenge Search, an attribute at a time iterative template search algorithm.
- Developed DFS algorithms to make use of MoLS's inferencing abilities.
- Developed the AK DFS algorithm which uses the AK heuristic which consistently provides savings in orders of magnitude.

Software Engineering a Framework

- Designed and prototyped the MoLS framework which allows the mining process to be customized using plug-ins based on a conceptual modularization of algorithms for searching powerset lattices.
- Built tools for inferencing as part of MoLS framework permitting developers to benefit from the ability to create savings without understanding how it is achieved.
- An iterative algorithm, Lozenge Search, allows the mining process to have sound and complete checkpoints at which AFD results can begin being used.
- The mining algorithm for a lattice is plugged into the Lozenge Search template and can be done so on a lattice by lattice basis.
- By building universal aspects of mining algorithms into the framework lattice search plug-ins become easier to write and code is reusable.

RESEARCH ASSISTANTSHIPS/PROJECTS

Data Integration as Concept Mapping, January 2008-May 2009

Associated Faculty: Rob Goldstone (Dept. of Cognitive Science Indiana University)

Graphs based on GIS data

- Induced a graph using Yahoo Local Search API where edges represent spatial relationships such as distance and bearing.
- Tested the robustness of ABSURDIST in handling imprecise data, i.e. bearing of Northwest instead of an exact bearing.
- Tested the subgraph matching problem wherein two graphs share some common subgraph or one graph entirely subsumes the other.

AbsMatcher for Schema Matching

- Induce a graph for attributes in a data set where edges represent relationships between attributes that are statistical in nature or represent semantic relatedness based on Yahoo! query results.
- Mined semantic similarity correspondences between attributes in two data sets.
- Utilized the ABSURDIST algorithm to find attribute-to-attribute mappings between two schemas, based on the mined graphs for each schema and semantic similarity correspondences.

Multiple Instance Learning in Chemistry Informatics, Summer 2007

Associated Faculty: Rajarshi Guha (Dept. of Informatics Indiana University)

- Implemented the basic EM algorithm combined with linear regression
- Used aggregate models to bootstrap the EM algorithm
- Improved the consistency of the algorithm by using noise to avoid local minimums

PROGRAMMING LANGUAGES

C, C++ STL, Java, Scheme, Visual Basic, HTML, Oracle PL/SQL, SQL Server, MYSQL