Katherine Grace Herbert
Assistant Professor
Department of Computer Science
Richardson Hall, Rm 318
Office: (973)655-5398
herbertk@mail.montclair.edu
http://www.csam.montclair.edu/~herbert/

EDUCATION

B.S., Computer Science, 1999, Saint Peter's College, Jersey City, New Jersey
B.S., Mathematics, 1999, Saint Peter's College, Jersey City, New Jersey
M.S., Computer Science, 2001, New Jersey Institute of Technology, Newark, New Jersey
Ph.D., Computer Science, 2004, New Jersey Institute of Technology, Newark, New Jersey

PROFESSIONAL EXPERIENCE

Montclair State University
Department of Computer Science
College of Sciences and Mathematics
Assistant Professor
Director, Biological Data Quality and Engineering Lab
Fall 2004 – Present

Data and Knowledge Engineering Lab, New Jersey Institute of Technology
Department of Computer Science, College of Computing Sciences
Research Associate, Summer 2004 - Present

Data and Knowledge Engineering Lab, New Jersey Institute of Technology
Department of Computer Science, College of Computing Sciences
Graduate Student Research Associate, Spring 2000 – May 2004

New Jersey Institute of Technology, Newark, New Jersey
Department of Computer Science, College of Computing Sciences
Teaching Assistantship, Fall 1999 – Spring 2004

AT&T, Florham Park, New Jersey
Research Intern - AT&T Labs Undergraduate Research Program, Summer 1997,1998

- 1998 - Researched database warehousing techniques
- 1997 - Researched feature interaction problems in telephony

HONORS AND AWARDS

AWARDS

- Mentor of Best Student Paper Award, ACM SIGCSE's 11th Annual Conference on Innovation and Technology in Computer Science Education (ITiCSE)
- Teaching Assistantship and Full Tuition award for PhD studies at New Jersey Institute of Technology (1999 – 2004)
- Summer 2003: Summer Research Assistantship Award from New Jersey Institute of Technology
- Summer 2003: Summer Research Assistantship Award from NSF Grant IIS-9988636
- Summer 2002: Summer Research Assistantship Award from NSF Grant IIS-9988636
- Summer 2001: Summer Research Assistantship Award from NSF Grant IIS-9988636
- Summer 2001: Summer Research Award: New Jersey I-TOWER Project
- May 1999: The Computer Science Award, Saint Peter's College, Jersey City, New Jersey

HONORS

- Who's Who Among America's Teachers (Student Nominated), 2004

TEACHING EXPERIENCE

Traditionally Taught Courses (Face to Face Learning):

Montclair State University

CSIT 110	Computer Concepts for Information Technology, Spring, Fall 2007, Spring 2008
CSIT 440	Data Mining, Fall 2006
CSIT 450	Text Management (using XML), Spring 2008
CMPT 109	Introduction to Computer Applications: Being Fluent with Information Technology, Summer, Fall 2005, Summer, Fall 2006, Spring 2007, Spring 2008
CMPT 183	Foundations of Computer Science I, Fall 2004, Spring 2006
CMPT 184	Foundations of Computer Science II, Spring 2005
CMPT 300	Introduction to Scientific Databases, Spring 2006, Spring 2008
CMPT 483	Database Systems, Fall 2007
CMPT 495	Special Topics in Computer Science: Data Mining, Fall 2006
CMPT 495	Special Topics in Computer Science: Introduction to Bioinformatics, Spring 2005
CMPT 495	Special Topics in Computer Science: Text Management: Perl and Bioinformatics Programming, Fall 2005
CMPT 585	Special Topics in Computer Science: Data Mining, Fall 2006

CMPT 585 Special Topics in Computer Science: Introduction to Bioinformatics, Spring 2005
CMPT 585 Special Topics in Computer Science: Text Management: Perl and Bioinformatics Programming, Fall 2005
CMPT 586 File Structures and Databases, Fall 2007
SCIF 152 Colloquium in Science Informatics II, Fall 2004
SCIF 253 Colloquium in Science Informatics III, Spring 2005
SCIF 491 Research Experience in Science Informatics I (with CMPT 495), Fall 2005, Fall 2006, Fall 2007
SCIF 492 Research Experience in Science Informatics II, Spring 2006, Spring 2007, Spring 2008

New Jersey Institute of Technology

CIS 113 Introduction to Computer Science I, Summer 2002, Fall 2002
CIS 114 Introduction to Computer Science II, Summer 2002, Summer 2004
CIS 350 Computers and Society, Summer 2000.
CIS 601 Object Oriented Programming, Summer 2004

Distance Learning Courses

New Jersey Institute of Technology

CIS 350 Computers and Society, Fall 1999, Spring 2000, Summer 2000, Fall 2000, Spring 2001, Summer 2001
CIS 431 Database System Design, Fall 2001

Courses Created

Montclair State University

CMPT 102 New Student Experience in Computers and Campus Society (co-authored with Angel Gutierrez), Spring 2007.
CSIT 450 Text Management, created Fall 2004, offered under CMPT 495/585 in Fall 2005.
CMPT 495/585 Special Topics in Computer Science: Introduction to Bioinformatics, Spring 2005
CMPT 585 Special Topics in Computer Science: Text Management: Perl and Bioinformatics Programming, Fall 2005

New Jersey Institute of Technology

CMPT 744 Data Mining and Management in Bioinformatics, created Summer 2003. Distance learning materials created with Jason T.L. Wang. Course has been offered during Fall 2004, Fall 2005 and Fall 2005 semesters at NJIT.

STUDENT RESEARCH AFFILIATIONS and PROJECTS

- Shekerah Primus and Frantz FIls-Aime, “Creating a database to assist in the statistical analyses of biological research data”, completed May 2008
- Brian Dugdale, “Biological Data Warehousing”, project under investigation as Master’s Project, completed May 2008
- Omer Metin Opaydin. “Applying Exploratory Data Mining Techniques on RNA Data”, project under investigation as master’s Project
- Jonathan Marra, “Sensitivity of Phylogenetic Tools for Gene Nomenclature Research”, completed Summer 2007
- Virginia L. Iourno “Biomolecular Data Mining and Knowledge Discovery Applied to Enzyme Modulation”, Master’s Project completed Spring 2007
- Jonathan Marra - “Automated Gene Sequence Retrieval System” Undergraduate Research Project completed Summer 2006
- Tazeen Fatima, Jonathan Marra, Ronald Realubit and Georgiy N. Shchegolev, “Automated Gene Sequence Retrieval System”, Science Informatics Undergraduate Research Experienece, Fall 2005 – Spring 2006.
- Georgiy N. Shchegolev, “Using Self Organizing Maps in Phylogenetic Data Analysis”, Undergraduate Research Project completed Summer 2005
- Tazeen Fatima, “Consensus Trees in Phylogenetics”, Undergraduate research Project completed Spring 2005

PEER REVIEWED PROCEEDINGS ARTICLES

1. Robert Barton, Sen Zhang, and Katherine G. Herbert, “Algorithms and Software for Calculating and Visualizing the Cardinality of the GDB”, in the Proceedings of The 2008 Internation Conference on Data Mining, July 2008, Las Vegas, Nevada, to appear.

2. Robert M. Siegfried, David M. Chays, and Katherine G. Herbert, “Will There Ever Be Consensus on CS1?”, in the Proceedings of The 2008 International Conference on Frontiers in Education: Computer Science and Computer Engineering, July 2008, Las Vegas, Nevada, to appear.

3. Jason T.L. Wang, Dongrong Wen, Bruce Shapiro, Katherine G. Herbert, Jing Li, Kaushik Gosh, “Toward an Integrated RNA Motif Database”, in the *Proceedings of the 4th International Workshop of Data Integration in the Life Science Workshop, Lecture Notes in Computer Science: SL 8 Bioinformatics*, June 2007, Philadelphia, Pennsylvania, pp 27-36.

4. Xiaoming Wu, Katherine G. Herbert, Jason T. L. Wang. “A New Kernel Method for RNA Classification”, in the *Proceedings of the IEEE 6th Symposium on Bioinformatics and Bioengineering*, October 2006, Arlington, Virginia, pp 201-208.

5. Sen Zhang, Katherine G. Herbert, Jason T.L. Wang, William H. Piel and David R.B. Stockwell. “PhyloMiner: a tool for evolutionary data analysis”, in the *Proceedings of 18th International Conference on Scientific and Statistical Database Management*, July, 2006, Vienna, Austria, pp 127 – 132.

6. Dorothy Deremer and Katherine G. Herbert. "An Interdisciplinary Undergraduate Science Informatics Degree in a Liberal Arts Context", in the *Proceedings of the 37th Technical Symposium on Computer Science Education, Houston, Texas, March 2006.*

7. Katherine G. Herbert and Jason T.L. Wang. "Phylogenetic Information Integration: Research Issues and Techniques", *Proceedings of the Joint Conference on Information Sciences 6th International Symposium on Computational Biology and Genome Information Systems & Technology*, Salt Lake City, Utah, July 2005, electronically published.

8. Katherine G. Herbert, Shashikanth Pusapati, Jason T. L. Wang, and William H. Piel." Lineage Path Integration for Phylogenetic Resources" *the Proceedings of 17th International Conference on Scientific and Statistical Database Management*, June, 2005, Santa Barbara, California, pp 117-120.

9. Jianghui Liu, Jason T.L. Wang, Wynne Hsu, and Katherine G. Herbert. "XML Clustering by Principle Component Analysis", in the *Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence,* November 2004, Boca Raton, Florida, pp 658-662.

10. Katherine G. Herbert, John Westbrook and Jason T.L. Wang , "Data Integration in Biological Databases", in the *Proceedings of the Joint Conference on Information Sciences 4th International Symposium on Computational Biology and Genome Information Systems & Technology*, Durham, North Carolina, September, 2003, pp 895-898.

11. Huiyuan Shan, Katherine G. Herbert, William Piel, Dennis Shasha, Jason T.L. Wang, "A Structure-Based Search Engine for Phylogenetic Databases", in *the Proceedings of 14th International Conference on Scientific and Statistical Database Management*, July, 2002, Edinburgh, Scotland, pp 7-10.

12. Katherine G. Herbert, Huiyuan Shan and Jason T.L. Wang "Approximate Searching in Phylogenetic Databases," *in the Proceedings of the Atlantic Symposium on Computational Biology and Genome Information Systems & Technology*, Durham, North Carolina, March 2001, pp. 140-143.

PEER REVIEWED JOURNAL ARTICLES

1. Katherine G. Herbert and Jason T.L. Wang. "Biological data cleaning: a case study", the International Journal of Information Quality, Vol 1, Issue 1, June 2007, pp 60-82.

2. Katherine G. Herbert and Dorothy Deremer. " Biological research through science informatics", CUR Quarterly, June 2006, pp 177-181.

3. Katherine G. Herbert, Narain H. Gehani, William H. Piel, Jason T.L. Wang, and Cathy H. Wu. "BIO-AJAX: An Extensible Framework for Biological Data Cleaning". *ACM SIGMOD Record, Special Issue on Data Engineering for the Life Sciences*, June 2004, pp51-57

4. Sen Zhang, Jason T.L. Wang and Katherine G. Herbert. "XML Query by Example," in *the International Journal of Computational Intelligence and Applications Special Issue on Internet Intelligence Systems*. World Scientific Publishing, Vol 2, No. 3 September 2002. Pages 329-338.

PEER REVIEWED CHAPTERS PUBLISHED

1. Katherine G. Herbert, Junilda Spirollari, Jason T.L. Wang, William H. Piel, John Westbrook, Winona Barker, Zhang-Zhi Hu, Cathy H. Wu, “Biological Databases”, The Encyclopedia of Computer Science and Engineering, Wiley & Sons, accepted.

2. Katherine G. Herbert, Jason T.L. Wang and Jianghui Lui. "Information Retrieval and Data Mining", *The Computer Science and Engineering Handbook*, Second Edition (ed. A. Tucker), CRC Press, June 2004, pp. 75-1 to 75-16.

3. Jason T. L. Wang, Qicheng Ma and Katherine G. Herbert, ``Software Engineering and Knowledge Engineering Issues in Bioinformatics,'' in *the Handbook of Software Engineering and Knowledge Engineering, Vol. 1, Fundamentals*, (ed. S. K. Chang), Chapter 30, World Scientific Publishing Company, 2001, pp. 719-732.

SOFTWARE AUTHORED

1. “Automated Gene Sequence Retrieval System”, advised students Tazeen Fatima, Jonathan Marra, Ronald Realubit and Georgiy N. Shchegolev in creation, Fall 2005 – Summer 2006.

2. “BIO-AJAX for Lineage Paths”, Dissertation research. Authored by Katherine G. Herbert, May 2004, significantly revised, January 2005.

3. “BIO-AJAX for TreeBASE”, Dissertation research. Authored by Katherine G. Herbert, November 2003.

4. **“**An automated annotation tool for courseware development”, Courseware on Demand Project. Authored by Vincent Oria, Katherine Herbert and Viswanath Neelavalli, for NJ-ITOWER project, December 2001.

TECHNICAL REPORTS AND POSTERS

1. Virginia L. Iuorno, Katherine G. Herbert, Jeffrey H. Toney, “A Support Vector Machine Method to Classify Enzyme Modulators”, in the Proceedings of the ISCB 5th Annual Rocky Mountain Bioinformatics Conference”, Snowmass, Colorado, December 2007 (faculty mentor).

2. Jason Caronna, Rojita Sharma, Jonathan Marra, Virginia L. Iuorno, Katherine G. Herbert and Jeffrey H. Toney, “Prediction of Modulators of Pyruvate Kinase in SMILES Text Using Aprori Methods”, in the Proceedings of the *12th Annual ACM SIGCSE International Conference on Innovation and Technology in Computer Science Education*, Dundee, Scotland, June 2007, pp 348 (faculty mentor for student poster)

3. Jonathan D. Marra, Katherine G. Herbert and Jason T.L. Wang. “A Study of Phylogenetic Tools for Genomic Nomenclature Data Cleaning”, in the Proceedings of the *12th Annual ACM SIGCSE International Conference on Innovation and Technology in Computer Science Education*, Dundee, Scotland, June 2007, pp 347 (faculty mentor for student poster)

4. Tazeen Fatima, Jonathan Marra*, Ronald Realubit, Georgiy Shchegolev and Katherine G. Herbert (faculty mentor). “Automated Gene Processing and Exon Sequence Retrieval”, in the Proceedings of the *11th Annual ACM SIGCSE International Conference on Innovation and Technology in Computer Science Education*, Bologna, Italy, June 2006, pp 366. Paper was a winner of a best student paper award at the conference.

5. Jason T.L. Wang, William H. Piel and Katherine G. Herbert. “What are the major informatics research challenges in biomedicine today?”, Request for Information Report to the National Library of Medicine, February 2006.

6. Katherine G. Herbert and James H. Dyer, “Science Informatics at Montclair State University”, *DIMACS Conference on Linking Mathematics and Biology in the High Schools*, Rutgers University, New Brunswick, New Jersey, April 29, 2005 (poster).

7. Katherine G. Herbert, “Biological Data Quality Research in the Department of Computer Science”, College of Science and Mathematics Newsletter, Montclair State University, Spring 2005.

8. Vincent Oria, Katherine G. Herbert, Viswanath Neelavalli, "An Automated Tool for Metadata Generation for Courseware-on-Demand", Technical Report submitted to NJ-ITOWER, December 2001.

PROFESSIONAL PRESENTATIONS/ABSTRACTS

1. “Integrating Interdisciplinary Science into High School Science”, presented with James H. Dyer at the 33rd Annual Trenton Computer Festival, Trenton, New Jersey, April 26, 2008.

2. “A Support Vector Machine Method to Classify Enzyme Modulators”, presented at the ISCB 5th Annual Rocky Mountain Bioinformatics Conference”, Snowmass, Colorado, December 1, 2007.

3. “Insights into Computer Science Graduate Studies”, presented for the Montclair State University NSF iMagine REU students, July 11, 2007.

4. “Toward an Integrated RNA Motif Database”, presented at the 4th International Workshop of Data Integration in the Life Science Workshop, Lecture Notes in Computer June 27, 2007, Philadelphia, Pennsylvania.

5. “Bioinformatics Activities in Computer Science”, Outreach Workshop for High School students participating in an admission day at Montclair State University, December 8, 2006.

6. “A New Kernel Method for RNA Classification”, presented at the *IEEE 6th Symposium on Bioinformatics and Bioengineering*, October 14, 2006, Arlington, Virginia.

7. “PhyloMiner: a tool for evolutionary data analysis”, presented at the *18th International Conference on Scientific and Statistical Database Management*, July 3, 2006, Vienna, Austria.

8. “What is Science Informatics?” with Dorothy Deremer, Charles Du, James H. Dyer and Aihua Li, the 2005 New Jersey Science Convention, Garden State Exhibition Center, Somerset, NJ, October 5, 2005.

7. “Developing a New Jersey Science Informatics Curriculum at Montclair State University” with Dorothy Deremer, Charles Du, James H. Dyer and Aihua Li, the 2005 New Jersey Science Convention, Garden State Exhibition Center, Somerset, NJ, October 5, 2005.

9. “Phylogenetic Information Integration: Research Issues and Techniques”, *Proceedings of the Joint Conference on Information Sciences 6th International Symposium on Computational Biology and Genome Information Systems & Technology*, Salt Lake City, Utah, July 25, 2005

10. “Science Informatics at Montclair State University” with James H. Dyer, presented at the *DIMACS Conference on Linking Mathematics and Biology in the High Schools*, Rutgers University, New Brunswick, New Jersey, April 29, 2005.

11. “Evolutionary Data Analysis and Data Management Issues”, invited presentation given at the *Saint Peter’s College Pi Mu Epsilon* annual meeting, Jersey City, New Jersey, April 25, 2005.

12. “XML Clustering by Principle Component Analysis”, presented at the IEEE *16th International Conference on Tools in Artificial Intelligence*, Boca Raton, Florida, USA, November 17, 2004.

13. “Data Cleansing and Knowledge Bases”, presented at the *4th Emerging Information Technologies Conference*”, Princeton, New Jersey, USA, October 30, 2004.

14. “Data Integration in Biological Databases", Presented at the *3rd Atlantic Symposium on Computational Biology and Genome Information Systems and Technology*, part of the 7th Joint Conference on Information Sciences, Raleigh, North Carolina, USA, September 26, 2003.

15. "TreeRank: A Similarity Measure for Nearest Neighbor Searching in Phylogenetic Databases", Presented at the *IEEE 15th International Conference on Scientific and Statistical Database Management*, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, July 10, 2003.

16. "A Structure-Based Search Engine for Phylogenetic Databases", Presented at the *IEEE 14th International Conference on Scientific and Statistical Database Management*, July 24 2002, University of Edinburgh, Edinburgh, Scotland.

17. "ATreeGrep: Approximate Searching in Unordered Trees", Presented at the *IEEE 14th International Conference on Scientific and Statistical Database Management*, July 24, 2002, University of Edinburgh, Edinburgh, Scotland.

18. "Approximate Searching in Phylogenetic Databases", Presented at the Atlantic Symposium on Computational Biology and Genome Information Systems & Technology, Durham, North Carolina, March 16, 2001.

MULTIMEDIA DISTANCE LEARNING COURSE MATERIAL

1. Katherine G. Herbert and Jason T.L. Wang "CIS 744: Data Mining and Management in Bioinformatics : CD 1", New Jersey Institute of Technology Instructional Technologies and Media Services, August 2003.

PROFESSIONAL SERVICE ACTIVITIES

University-Based Service:

University
- University Parking Review Committee (2004-1005)

College
- Department Specialist in Bioinformatics for the Science Informatics Coordinating Committee: January 2006 – present
- Member, Science Informatics Coordinating Committee, Fall 2004- December 2005
- Advisor to Science Informatics Students, Concentration in Computer Science, Fall 2004- December 2005

Department
- Assessment Committees for CMPT 109 and CMPT 183
- CMPT 109 Challenge Exam Development Committee
- Steering Committee for Proposed Ph.D. Program in Computational Sciences
- Advisor for Computer Science Major

Discipline Based Service

Panel Reviews:
- National Science Foundation Division of Undergraduate Education S-STEM-08 (January 2008)

Conference and Journal Peer Publication Reviews:

- Knowledge and Information Systems Journal (KAIS 2005, 2006, 2007, 2008)
- The 2008 IEEE World Congress on Computational Intelligence
- International Journal of Data Mining and Bioinformatics (February, March 2007, April 2008)
- 17th European Conference on Machine Learning and the 10th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD 2006)
- International Association of Science and Technology for Development Conference 2006
- The VLDB (Very Large Database) Journal (2005)
- The IEEE Fifth International Conference on Data Mining (ICDM 2005)
- The 28th Australasian Computer Science Conference 2005 (ACSC 2005)
- VLDB Journal, Fall 2004
- Information Sciences (Journal Published by Elsevier Publishing), Fall 2004
- 28th Australasian Computer Science Conference-2005, (ACSC2005)
- 2004 SIAM International Conference on Data Mining
- The Internet Encyclopedia, 2004
- The 14th International Conference for Intelligent Systems for Molecular Biology (ISMB 2003)
- Information Systems: Special Issue on Bioinformatics and Biological Data Management, June 2003
- The International Conference on Information Technology: Research and Education (ITRE 2003), August 2003
- The International Conference on Intelligent Systems for Molecular Biology, July 2003
- The 3rd SIAM International Conference on Data Mining, May 2003
- The IEEE International Conference on Data Mining, December 2002
- The IEEE 4th International Symposium on Multimedia Software Engineering, December 2002
- The International Journal on Computational Intelligence and Applications: Special Issue on Intelligent Systems, September 2002
- The 3rd International Conference on Web-Age Information Management, August 2002
- The 6th Pacific-Asia Conference on Knowledge Discovery and Data Mining, May 2002
- The 2nd SIAM International Conference on Data Mining, April 2002
- The 14th International FLAIRS Conference, May 2001
- The IEEE International Conference on Data Mining, December 2001

Conference Committee Participation and Conference Session Chair Service:

- Technical Committee Member, The 2008 IEEE World Congress on Computation Intelligence
- Program Committee Member, BIOCOMP'07- The 2007 International Conference on Bioinformatics & Computational Biology

- Session Chair, BIBE 2006, Arlington, Virginia, 2006
- Program Booklet Committee, ACM SIGMOD, June 2005
- Session Chair, CBGI 2005, Salt Lake City Utah, July 2005
- Session Chair, CCSC 2004, Baltimore, Maryland, October 2004
- Session Chair, CBGI 2003, Raleigh, North Carolina, 2003

ASSOCIATION MEMBERSHIPS

Discipline-based memberships:

- Association for Computing Machinery (ACM)
- The International Society of Computational Biology (ISCB)
- Institute for Electrical and Electronic Engineers (IEEE)

Honor Societies

Inducted to these societies as a student and have maintained membership

- Alpha Sigma Nu (The National Jesuit Honors Society – Spring 1997)
- Pi Mu Epsilon (The National Mathematics Honors Society – Spring 1997)
- Upsilon Pi Epsilon (The ACM International Computer Science Honors Society – Fall 2002)

REFERNENCES

Furnished upon request