

# Doina Caragea

## Contact Information

2184 Engineering Hall  
Department of Computer Science  
Kansas State University  
Manhattan, KS 66506 USA

*Voice:* 785-532-7908  
*Fax:* 785-532-7353  
*E-mail:* dcaragea@ksu.edu  
*WWW:* www.cs.ksu.edu/~dcaragea

## Research Interests

Big Data, Machine Learning, Data Mining, Data Science, Text Mining, Social Network Analysis, Security Informatics, Recommender Systems, Bioinformatics.

## Education

- **Ph.D., Iowa State University, Computer Science, Ames, IA, USA** *August 2004*
  - Thesis Title: “Learning Classifiers from Distributed, Semantically Heterogeneous, Autonomous Data Sources.”
  - Minor: Bioinformatics
  - Advisor: Vasant Honavar
- **M.S., University of Bucharest, Computer Science, Bucharest, Romania** *July 1997*
  - Thesis Title: “Statistical Genetic Algorithms.”
  - Advisor: Luminita State
- **B.S., University of Bucharest, Computer Science, Bucharest, Romania** *July 1996*

## Professional Experience

- **Associate Professor** *July 2012 - present*  
*Department of Computer Science, Kansas State University*
  - Projects: domain adaptation and semi-supervised learning with application to crisis-related classification, Android malware detection, recommender systems, bioinformatics, etc.
- **Assistant Professor** *August 2006 - July 2012*  
*Department of Computer Science, Kansas State University*
  - Projects: machine learning applications to bioinformatics, genomic sequence analysis, knowledge discovery from semantically heterogeneous data, etc.
- **Associate Director** *January 2007 - present*  
*Bioinformatics Center, Kansas State University*
  - Projects: Gene regulatory networks, EST sequence analysis, NGS sequence analysis, gene prediction, genome annotation, etc.
- **Postdoctoral Research Associate** *August 2004 - August 2006*  
*Department of Computer Science, Iowa State University*

- Project: Information integration and knowledge acquisition from autonomous, distributed, semantically heterogeneous data sources.
- Advisor: Vasant Honavar

- **Graduate Research Assistant**

*August 1998 - July 2004*

- *Department of Computer Science, Iowa State University*

- Projects: Incremental and Distributed Learning Algorithms, Sufficient Statistics for Machine Learning Algorithms, Information Integration and Ontologies, Visual Data Mining, etc.
- Advisor: Vasant Honavar

- **Summer Intern**

*May - August 2003*

- *IBM Rochester, MN*

- Project: MineLink - A Federated Information Integration Framework Using Automatic Service Composition
- Project leader: Tanveer Syeda-Mahmood (IBM Almaden)
- Manager: Lance Snow

- **Summer Intern**

*May - August 2002*

- *IBM Rochester, MN*

- Project: ABLE - Agents Based Learning Environment
- Project leader: Joe Bigus (IBM T.J. Watson)
- Manager: Sam Ellis

- **Research Associate**

*July 1996 - July 1998*

- *Institute of Microtechnology, AI Group, Bucharest, Romania*

- Project: GA vs. Statistical GA vs. Messy GA, Evolvable Hardware with Evolutionary Algorithms and Reinforcement Learning, etc.
- Manager: Florin Fagarasan

## Honors and Awards

- Best paper award, IEEE/ACM International Symposium on Network Enabled Health Informatics, Biomedicine and Bioinformatics (HI-BI-BI 2016), *2016*
- Best paper award, IEEE/ACM International Symposium on Network Enabled Health Informatics, Biomedicine and Bioinformatics (HI-BI-BI 2015), *2015*
- Best paper award finalist, The 5th International Conference on Bioinformatics Models, Methods and Algorithms (BIOINFORMATICS 2014), *2014*
- Best student paper award finalist, The 4th International Conference on Bioinformatics Models, Methods and Algorithms (BIOINFORMATICS 2013), *2013*
- Best paper award finalist, The 10th International Conference on Machine Learning and Data Mining (MLDM 2013), *2013*
- Best paper award, The First Asian Semantic Web Conference (ASWC 2006), *2006*
- Iowa State University *Research Excellence Award*. Selected by The Office of the Graduate Dean and Vice Provost for Research and Advanced Studies, *August 2004*
- IBM Fellowship (awarded to 54 Ph.D. students worldwide), *2003-2004*

- IBM Fellowship (awarded to 56 Ph.D. students in the US), 2002-2003

## Research Grants

1. *BIGDATA: IA: Collaborative Research: Domain Adaptation Approaches for Classifying Crisis Related Data on Social Media*. National Science Foundation. **Doina Caragea** (PI - Kansas State University with co-PIs Cornelia Caragea and Dan Andresen), Andrea Tapia (Pennsylvania State University, with co-PI Jess Kropczynski). Award amount: \$1,300,000, KSU amount: \$900,000, 2018-2022.
2. *SaTC: CORE: Small: Collaborative: Data-driven Approaches for Large-scale Security Analysis of Mobile Applications*. National Science Foundation. **Doina Caragea** (PI-Kansas State University), Xingmin Ou (PI-University of South Florida), Sankardas Roy (PI-Bowling Green State University). Award amount: \$500,000, KSU award amount: \$200,000, 2017-2020.
3. *MRI: Acquisition of an Adaptive Data Cluster for Data-Intensive Applications in Science and Engineering*. National Science Foundation, Dan Andresen (PI), **Doina Caragea** (co-PI), Susan Brown (co-PI), Xinming Ou (co-PI), Jesse Poland (co-PI). Award amount: \$467,000, 2014-2016.
4. *CC-IEE Networking Infrastructure: KGEN: Next-generation Networking Environments for Biological and Agricultural Data-driven Research at Kansas State University*. National Science Foundation. **Doina Caragea** (senior personnel), with Dan Andresen, PI). Award amount: \$231,500, 2014-2016.
5. *CC-NIE Network Infrastructure: KGAP: Bridging the gap in network flexibility and performance for genomics and data-intensive research at Kansas State University*. National Science Foundation, Dan Andresen (PI), **Doina Caragea** (co-PI), J.M. Shawn Hutchinson (co-PI), Don Gruenbacher (co-PI), Susan Brown (co-PI). Award amount: \$499,113.00, 2013-2016.
6. *MRI: Acquisition of a Hybrid GPU Computing Cluster High-End Applications in Science and Engineering*. National Science Foundation. Dan Andresen (PI), **Doina Caragea** (co-PI), David Steward (co-PI), Walter Dodds (co-PI), Brett Esry (co-PI). Award amount: \$840,000, 2012-2015.
7. *Comparative Genomics of Flour Beetles in the Genus Tribolium*. Seed Grant Arthropod Genomics Center, KSU, Susan Brown (PI), **Doina Caragea** (co-PI), Richard Beeman (co-PI), Yoonseong Park (co-PI) and Subbarantnam Muthukrishnan (co-PI). Award amount: \$120,000, 2010-2012.
8. *Comparative transcriptome sequence analysis of two host races of the Grasshopper *Hesperotettix viridis* - Searching for evidence of host associated divergence and incipient speciation*. Seed Grant, Integrated Genomics Facility, KSU, Tony Grace (PI), Susan Brown (co-PI), **Doina Caragea** (co-PI), Samantha Wisely (co-PI) and Anthony Joern (co-PI). Award amount: \$10,000, 2010-2011.
9. *Cyberinfrastructure Implementation for Genotype to Phenotype Research*. National Science Foundation. Steve Welch (PI), **Doina Caragea** (co-PI) and Sanjoy Das (co-PI). Award amount: \$314,847, 2009-2011.
10. *TRMS: Ecological Annotation of Gene Function and Computational Analysis of Gene Networks*. National Science Foundation, DBI - Plant Genome Research Project. **Doina Caragea**, senior personnel (with Cynthia Weinig PI and Steve Welch, Justin Maloof, Sanjoy Das co-PIs). Award amount: \$5,652,782, KSU award amount: \$1,212,620, 2010-2015.

11. *EPSCoR TRACK II Oklahoma and Kansas: Wiring the Central Plains: Cyberinfrastructure to Monitor and Model Ecosystems Under Directional Change*, National Science Foundation. **Doina Caragea** senior personnel (with Walter Dodds PI et al.). KSU award amount: \$1,605,472, 2009-2012.
12. *Computational Methods to Characterize Regulatory Networks Involved in Plant Response to Abiotic Stresses*. KSU Ecological Genomics Seed Grant. Haiyan Wang (PI), **Doina Caragea** (co-PI) and Susan J. Brown (co-PI). Award amount: \$35,589, 2008-2009.
13. *Advanced Genomics at K-State: Ultra-High Throughput DNA Sequencing. KSU Targeted Excellence Program*. Eduard Akhunov, Bikram Gill, Frank White, Karen Garrett, James Nelson, Susan Brown, Loretta Johnson, Michael Herman, Jianming Yu, and Sanjeev Narayanan, Ludek Zurek and **Doina Caragea** as co-principal investigators. Award amount: \$850,000, 2008-2011.
14. *Collaborative Research: Learning Classifiers from Autonomous, Semantically Heterogeneous, Distributed Data*. National Science Foundation. **Doina Caragea** (PI-KSU), Vasant Honavar (PI-ISU). Award amount: \$449,999, KSU award amount: \$145,504, 2007-2010.
15. *IIS-0639230 SGER: Exploratory Investigation of Modular Ontologies*. National Science Foundation. Vasant Honavar (PI), Giora Slutzki (Co-PI), and **Doina Caragea** (Co-PI). Award amount: \$112,000, 2006-2008.

### Professional Development and Travel Awards

- *ADVANCE Distinguished Lecture Speaker Award*, \$1200. Invited Dr. Raghu Ramakrishnan from University of Wisconsin, Madison to give a lecture on “The Future of Information Discovery: Content Optimization, Social Networks, Interactivity - It’s All Very Cloudy,” *Spring 2012*
- *ADVANCE Research Enhancement Visit Award*, \$1032. Travel Grant to IBM Watson, *Fall 2008*
- *ADVANCE Distinguished Lecture Speaker Award*, \$1200. Invited Dr. Michael Lynch from Indiana University to give a lecture on “The Origins of Genome Architecture,” *Spring 2008*
- *KSU President’s Faculty Development Award*, \$2500. Travel Grant to the International Conference on Intelligent Systems for Molecular Biology, ISMB 2007, Vienna, Austria, *Fall 2007*
- *ADVANCE Research Enhancement Visit Award*, \$802. Travel Grant to NSF, *Fall 2007*
- *ADVANCE Distinguished Lecture Speaker Award*, \$1200. Invited Dr. Volker Brendel from Iowa State University to give a lecture on “Necessity and Promise of Automated and Community Genome Annotation,” *Fall 2007*
- *KSU President’s Faculty Development Award*, \$2700. Travel Grant to the Int. Joint Conference on Artificial Intelligence (IJCAI’07), Hyderabad, India, *Spring 2007*
- *KSU Arthropod Genomics Center Award*, \$1000. Travel Grant to RECOMB 2007, *Spring 2007*

## Teaching

- **CIS 560:** *Database System Concepts* (3 credits) Undergraduate level course, which introduces concepts, approaches, and techniques in database management. The topics covered include exploring the representation of information as data, data storage techniques, foundations of logical data models, data retrieval, database design, transaction management, integrity and security. I started to teach this course in Fall 2010, and I have taught it every fall semester since then (except for Fall 2012). I completely redesigned the course with Stephane Faroult in Fall 2015, and I have taught it together with CIS562 (previously taught by Stephane) in Fall 2016 and Fall 2017 (roughly 100 students each year).
- **CIS734/Biol734:** *Introduction to Genomics and Bioinformatics* (4 credits) Undergraduate and graduate level course that I originally designed and developed, together with Dr. Susan J. Brown in the Division of Biology, in Spring 2007. The course provides fundamental background in bioinformatics, both theoretical (bioinformatics algorithms) and practical (databases and web-based tools used to study problems in biology), to students in computer science or in biological sciences. The course has been continuously updated to track the new developments in Next Generation Sequence (NGS) analysis. Practical use of the tools introduced in class is demonstrated by laboratory exercises and homework problems. Students are grouped for class assignments, such that each group contains at least one life scientist and one computer scientist. Course projects have been generally related to ongoing biological research projects at KSU. I have taught this course in Spring 2007-2010, Spring 2012, Spring 2014, and Spring 2016.
- **CIS 761:** *Database Management Systems* (3 credits) Graduate level course that I designed in Spring 2013, and taught every spring semester since then. The purpose of this course is to introduce concepts, approaches, and techniques in database management. This includes exploring the representation of information as data, data storage techniques, foundations of logical data models, data retrieval, database design, transaction management, integrity and security. It also includes topics such as NoSQL, BigData, MapReduce, Spark.
- **CIS 833:** *Information Retrieval and Text Mining* (3 credits) Graduate level course that I designed, and taught in Fall 2008-2011, Spring 2013, Fall 2013-2016. This course is focused on the theory and practice of search engines for retrieving textual information (including web documents). Basic and advanced topics in information retrieval (IR) are covered, with emphasis on newer technologies that go beyond simple keyword search. The topics targeted in this course include: efficient text indexing, latent semantic indexing, Boolean and vector space retrieval models, probabilistic retrieval models, language models, evaluation, web spiders, link analysis (PageRank and HITS), relevance feedback and query expansion, text mining and categorization, topic detection and clustering, etc. Given the need for processing large amounts of text data in information retrieval, the course will also cover MapReduce basics and algorithm design, and more recently Spark. Programming assignments in Hadoop or Spark provide hands-on experience with retrieval systems and big data technologies. More advanced research in IR is stimulated through the means of a class project.
- **CIS 890:** *Topics in Deep Learning* (3 credits) This is a discussion-based course targeted at graduate students in computer science and related areas, which I am offering for the first time in Fall 2017. The course covers the foundations of deep learning, with focus on different types of modern neural network (such as convolutional neural networks, recurrent and recursive neural networks, long short-term memory networks, autoencoders). Furthermore, the course also covers tools and software available for building and training deep neural networks. Basic knowledge in machine learning is expected.

- **CIS 890: *Machine Learning for Bioinformatics*** (3 credits) Graduate level course that I designed and taught, with assistance from Dr. Susan J. Brown in the Division of Biology, in Fall 2007, and taught again in Spring 2011. In this course, some of the most important machine learning algorithms and their applications to bioinformatics are discussed. Machine learning topics include: probabilistic and graphical models, network models, support vector machines, etc. The bioinformatics applications are in the areas of structural and functional genomics and proteomics. Specific examples include: transcription binding site prediction, protein function, localization and structure prediction, protein-protein interaction prediction, etc.

## Teaching Enhancement

- Spring 2009 - Attended The Sixth Annual Teaching Retreat for K-State teaching faculty in January 2009. The retreat's theme was "On the road to teaching excellence: our journey as teachers." Organized by The Center and the Faculty Exchange for Teaching Excellence at K-State.
- Fall 2008 - Attended Dr. Maryellen Weimer's (Penn State University) workshop on "Grades, Assessment, and a Focus on Learning." The workshop was hosted by The Center and the Faculty Exchange for Teaching Excellence at K-State.
- Fall 2006, Spring 2007 - Participated in the KSU LEA/RN program (2006-2007), a faculty development program that was offered by the College of Engineering at K-State. LEA/RN focused on student-centered learning, active learning techniques, active learning lesson planning, and selected topics in learning theory.
- Fall 2006, Spring 2007 - Attended more than four sessions in the New Faculty Institute (NFI) and "graduated" with an award plaque and certificate. NFI facilitates the transition of newly-hired faculty members into the K-State academic environment by providing information and tools that can help them become successful as educators and researchers at K-State.

## Student Advising

### Current PhD Students

1. *HongMin Li*, expected graduation date: May 2019
2. *Chaoxin Wang*, expected graduation date: TBD
3. *Nawaf Alharbi*, expected graduation date: TBD

### Past PhD Students

1. *Nic Herndon* (Ph.D., 2016), Thesis: *Domain Adaptation Algorithms for Biological Sequence Classification*. Initial employment: Postdoctoral Fellow at University of Connecticut.
2. *Ana Stanescu* (Ph.D., 2015), Thesis: *Semi-supervised Learning for Biological Sequence Classification*. Initial employment: Postdoctoral Fellow at Icahn School of Medicine at Mount Sinai; currently, Assistant Professor in the Computer Science at University of West Georgia.

3. Rohit Parimi (Ph.D., 2015), Thesis: *Collaborative Filtering Approaches for Single-Domain and Cross-Domain Recommender Systems*. Initial employment: Senior Software Engineer at Motorola, Chicago, IL; currently, Senior Software Engineer at Bloomberg LP, New York, NY.
4. Karthik Tangirala (Ph.D., 2015), Thesis: *Unsupervised Feature Construction Approaches for Biological Sequence Classification*. Initial employment: Software Development Engineer II at Microsoft, Seattle, WA; currently, Data & Applied Scientist II at Microsoft, Seattle, WA.

### Current M.S./M.S.E. Students

1. Reza Mazloom, M.S. thesis, expected graduation date: May 2018.
2. David Ungurean (exchange student), M.S. thesis, expected graduation date: May 2018.
3. Emily Alfs, M.S. thesis, expected graduation date: May 2018.
4. Vijay Kumar Venkatamuniyappa, M.S. report, expected graduation date: May 2018.
5. Akash Rathore, M.S. report, expected graduation date: May 2018.
6. Zhiang Fan, M.S. report, expected graduation date: May 2018.
7. Xukun Li, expected graduation date: TBD

### Past M.S./M.S.E. Students

1. Chandrika Mitra (M.S.E., 2017), *Scientific Article Recommender System based on Collaborative Denoising Auto-Encoders*, M.S.E. project.
2. Bhavani Krithivasan (M.S., 2017), *Cross-language text classification*, M.S. report.
3. Oleksandra Sopova (M.S., 2017), *Domain adaptation for classifying disaster-related Twitter data*, M.S. report.
4. Tyler Robinson (M.S., 2016) *Disaster Tweet Classification using Parts-of-Speech Tags: A Domain Adaptation Approach*, M.S. thesis.
5. David McWaters (M.S.E., 2016) *Author-Paper Identification System Using Neural Networks*, M.S.E. project.
6. Tomas Trepka (M.S., 2015), *Ad-recommending system*, MS thesis.
7. Ankit Bajpai (M.S., 2014), *SQL Versus NoSQL from an Application Development Point of View*, M.S. report.
8. Heath Yates (M.S., 2014), *Study and comparison of next generation sequence algorithms and tools*, M.S. report.
9. Anirudh Jagithyala (M.S., 2014), *Recommending recipes based on ingredients and user reviews*, M.S. thesis.
10. Sunny Kumar (M.S.E., 2014), *Author Identification System*, M.S.E. project.
11. Paul Cain (M.S.E., 2013), *Queryable RNA-Seq Database*, M.S.E. project.
12. Matt Cholick (M.S.E., 2012), *Movie Tweets: A Micro-blogging Bootstrapped Recommender*, M.S.E. project.

13. Swapnil Nagar (M.S., 2012), *A hybrid recommender: user profiling from tags/keywords and ratings*, M.S. thesis.
14. Vishwas Vaswani (M.S., 2012), *Predicting sentiment-mention associations in product reviews*, M.S. thesis.
15. Srilaxmi Cheeti (M.S., 2012), *Cross-domain sentiment classification using grams derived from syntax trees and an adapted naive Bayes approach*, M.S. thesis.
16. Sandeep Solanki (M.S., 2012), *Engineering enhancements for movie recommender systems*, M.S. thesis.
17. Tri Doan (M.S., 2012), N/A.
18. Rahul Choubey (M.S., 2011), *Tag recommendation using Latent Dirichlet Allocation*, M.S. thesis.
19. Karthik Tangirala (M.S., 2011), *Semi-supervised and transductive learning algorithms for predicting alternative splicing events in genes*, M.S. thesis.
20. Surbhi Mungre (M.S., 2011), *LDA-based dimensionality reduction and domain adaptation with application to DNA sequence classification*, M.S. thesis.
21. Shruti Phanse (M.S., 2010), *Study on the performance of ontology based approaches to link prediction in social networks as the number of users increases*, M.S. thesis.
22. Roshan Chetry (M.S., 2011), *Web genre classification using feature selection and semi-supervised learning*, M.S. report.
23. Rohit Parimi (M.S., 2010), *LDA based approach for predicting friendship links in live journal social network*, M.S. thesis.
24. Vishal Bahirwani (M.S., 2010), *Exploring transcription patterns and regulatory motifs in Arabidopsis thaliana*, M.S. thesis.
25. Swarnim Kulkarni (M.S., 2009), *Capturing semantics using a link analysis based concept extractor approach*, M.S. thesis.
26. Mandar Haridas (M.S., 2009), *Exploring knowledge bases for engineering a user interests hierarchy for social network applications*, M.S. thesis.
27. Aditi Breed (M.S., 2008), *Querying semantically heterogeneous data sources using ontologies*, M.S. thesis.
28. Vikas Bahirwani (M.S., 2008, co-advised with Dr. William H. Hsu), *Ontology engineering and feature construction for predicting friendship links and users interests in the Live Journal social network*, M.S. thesis.
29. Martin Paradesi (M.S., 2008, co-advised with Dr. William H. Hsu), *Graph-based protein-protein interaction prediction in Saccharomyces cerevisiae*, M.S. thesis.

### Current Undergraduate Students

- Joosung Ko (undergraduate research assistant)
- Joe Malone (undergraduate research assistant)
- Matthew Link (undergraduate research assistant)



- Alice Lam (senior project)
- Zakary Kedrovsky (senior project)
- Matthew Hixon (senior project)

### Past Undergraduate Students

1. Daniel Longfellow (Fall 2017): Digest App for Twitter (Senior project)
2. Lowell Scott (Fall 2017): Comparative Statistics for Three Hurricanes (Senior project)
3. Jonathan Howard (Fall 2017): Disaster Event Notifier (Senior project)
4. Eujun Chin (Fall 2017): App for Finding Roommates (Senior project)
5. Jordan DeLoach (2015-2017): Machine Learning for Android Security (undergraduate research assistant, senior project). Caragea nominated Jordan for the *Raj and Diana Nathan Undergraduate Research Experience Award* for Academic Year 2016 - 2017. Jordan was the recipient of the award and received \$5,000 to do research on “Android malware detection based on the wisdom of crowds” under D. Caragea’s supervision, *2016-2017*.
6. Steven Mercier (Spring 2017): Upgrading a Database from Access 2003 to PostgreSQL (Senior project)
7. Mohammed Alsayyari (Spring 2016): Name Disambiguation in Author Databases (Senior project)
8. Hayden Wilson (Fall 2015) - Notes taking app: speech to text (senior project)
9. Joao Carlos Ferreira Marques (Summer 2015) - Music Database: Cleaning Discogs Data (summer intern supported by the Brazil Scientific Mobility Program)
10. Vibhore Bhatnagar (Spring 2015) - Extracting Metadata from pdf Files (senior project, incomplete)
11. Trevin Garcia (2014-2015) - News Aggregation Approaches (DSP student)
12. Ana Natalia Donaldson (Summer 2014) - Sentiment Analysis of Twitter Data (summer DREU student from University of New Mexico funded by CRA-W)
13. Ryan Hershberger (Spring 2014) - Track and Field Database System (senior project)
14. Ryan Dry (Spring 2014) - Irys Database - web app for managing bionano data (senior project)
15. Geordy Williams (2013-2014) - Using Recommender Systems for Social Network Applications (DSP student)
16. Heather Carmitchel (Spring 2013): Media Ratings & Statistics Website (senior project)
17. Tanmay Varma (Spring 2013): Business Intelligence and Topic Modeling (senior project)
18. Ryan Schmidt (Fall 2012): Twitter Sentiment Analysis (senior project)
19. Katherine Mawhinney (Spring 2012) - Twitter Data Mining (senior project)
20. Robert Teutsch (Spring 2012) - Exploration of Bioinformatics Tools (undergraduate research assistant at the Bioinformatics Center)

### Visiting Scholars

1. Baohua Yang (Spring 2017) - Associate Professor, College of Information and Computer, Anhui Agricultural University, China.
2. Liu Chen (Spring 2015) - Lecturer, School of Electrical and Information Engineering, Wuhan Institute of Technology, China.

## Publications

### Ph.D. Thesis

**Caragea, D.** (2004). *Learning Classifiers from Distributed, Semantically Heterogeneous, Autonomous Data Sources*. Iowa State University.

### Refereed Journal Articles

1. Noll, L.W., Worley J.N., Yang, X., Shridhar, P.B., Ludwig, J.B., Shi, X., Bai, J., **Caragea, D.**, Meng, J., Nagaraja, T.G. (2018) *Comparative genomics reveals differences in mobile virulence genes of Escherichia coli O103 pathotypes of bovine fecal origin*. PLoS ONE 13(2): e0191362.
2. Li, H., **Caragea, D.**, Caragea, C. and Herndon, N. (2017). *Disaster Response Aided by Tweet Classification with a Domain Adaptation Approach*. In: Journal of Contingencies and Crisis Management (JCCM), Special Issue on HCI in Critical Systems. In press.
3. Noll, L.W., Worley J.N., Yang, X., Shridhar, P.B., Bai, J., Meng, J., **Caragea, D.**, and Nagaraja, N.G. (2017). *Draft Genome Sequences of Enterohemorrhagic Escherichia coli O103: H2 Strains Isolated from Feces of Feedlot Cattle*. In: Genome announcements, Vol. 5, Issue 19, pages e00094-17
4. Noll, L.W., Worley J.N., Yang, X., Shridhar, P.B., Bai, J., Meng, J., **Caragea, D.**, and Nagaraja, N.G. (2017). *Draft Genome Sequences of Enteropathogenic Escherichia coli O103 Strains Isolated from Feces of Feedlot Cattle*. In: Genome announcements, Vol. 5, Issue 21, pages e00387-17
5. Stanescu, A. and **Caragea, D.** (2017). *An empirical study of self-training and data balancing techniques for splice site prediction*. International Journal of Bioinformatics Research and Applications. Volume 13, No. 1, pp. 40 - 61, February 2017.
6. Neppalli, K., Caragea, C., **Caragea, D.**, Medeiros, M.C., Tapia, A. and Halse, S. (2016). *Predicting Tweet Retweetability during Hurricane Disasters*. International Journal of Information Systems for Crisis Response and Management (IJISCRAM). Vol. 3, no. 8, pages 32-50.
7. Herndon, N., and **Caragea, D.** (2016). *An evaluation of approaches for using unlabeled data with domain adaptation*. In: Network Modeling Analysis in Health Informatics and Bioinformatics. 5(25):1-12, 2016.
8. Tangirala, K., Herndon, N. and **Caragea, D.** (2016). *A Comparative Analysis between k-mers and Community Detection-based Features for the Task of Protein Classification*. In: Special Issue of Transactions on Nanobioscience. Vol. 15, Issue 2, pp. 84 - 92, March 2016.
9. Herndon, N. and **Caragea, D.** (2016). *A Study of Domain Adaptation Classifiers Derived from Logistic Regression for the Task of Splice Site Prediction*. In: Special Issue of Transactions on Nanobioscience. Vol. 15, Issue 2, pp. 75-83, March 2016.
10. Stanescu, A., Tangirala, K. and **Caragea, D.** (2016). *Predicting Alternatively Spliced Exons Using Semi-supervised Learning*. In: International Journal on Data Mining and Bioinformatics (IJDMB). Vol. 14, No. 1, pages 1-21.

11. Stanescu, A., and **Caragea, D.** (2015). *An empirical study of ensemble-based semi-supervised learning approaches for imbalanced splice site datasets*. In: BMC System Biology. Vol 9 (Suppl 5):S1.
12. Zhang, S., Ou, X., and **Caragea, D.** (2015). Predicting Cyber Risks through National Vulnerability Database, Information Security Journal: A Global Perspective. Vol. 24, Issue 4-6, 2015.
13. Zhao, C., Navarro Escalante, L., Benatti, T., Qu, J., Chellapilla, S., Waterhouse, R. M., . . . , **Caragea, D.**, . . . Richards, S. (in press). *A Massive Expansion of Effector Genes Underlies Gall-Formation in the Wheat Pest (Mayetiola destructor)*. Current Biology, 2015 Feb 1.
14. Carolan, J.C., **Caragea, D.**, Reardon, K.T., Mutti, N.S., Pappan, K., Dittmer, N, Cui, F., Reeck, G.R, Castaneto, M., Poulain, J., Dossat, C., Wilkinson, T.L., Tagu, D., Reese, J.C. and Edwards, O.R. *Predicted effector molecules in the salivary secretome of the pea aphid (Acyrtosiphon pisum) - a dual transcriptomic/proteomic approach*. Journal of Proteome Research, 2011 Apr 1;10(4):1505-18.
15. Caragea, C., **Caragea, D.**, Silvescu, A., and Honavar, V. *Semi-supervised prediction of protein subcellular localization using abstraction augmented Markov models*, Special Issue on Machine Learning in Computational Biology (MLCB), BMC Bioinformatics. 2010 Oct 26;11 Suppl 8:S6.
16. Steller, M., Kambhampati, S., and **Caragea, D.** *Comparative analysis of expressed sequence tags from three castes and two life stages of the termite Reticulitermes flavipes*, BMC Genomics 2010, 11:463.
17. Travers, S. E., Tang, Z., **Caragea, D.**, Garrett, K.A., Hulbert, S. H. , Leach, J. E. , Bai , J., Saleh, A., Knapp, A.K., Fay, P.A., Nippert, J., Schnable, P.S., and Smith, M.D. *Spatial and temporal variation of big bluestem (Andropogon gerardii) transcription profiles with climate change*. In: Journal of Ecology, 2010, 98, 374-383. *Selected as Editor's Choice for March 2010*.
18. Kim, H.S., Murphy, T., Xia, J., **Caragea, D.**, Park, Y., Beeman, R., Lorenzen, M., Manak, J., Butcher, S. and Brown, S. *BeetleBase in 2010: revisions to provide comprehensive genomic information for Tribolium castaneum*. In: Journal of Nucleic Acid Research (NAR), 2010, Vol. 38, Database issue D437-D442.
19. Xia, J., **Caragea, D.** and Brown, S.J. *Prediction of alternatively spliced exons using support vector machines*. In: International Journal on Data Mining and Bioinformatics (IJDMB). Vol. 4, No. 4, 411-430, 2010.
20. **Caragea, D.**, Silvescu, A., and Honavar, V. *A Framework for learning from distributed data using sufficient statistics and its application to learning decision trees*. Invited Paper. In: International Journal of Hybrid Intelligent Systems. Vol. 1, No. 2, Pp. 80-89, 2004.

## **Books**

1. Das, S., **Caragea, D.**, Hsu, W. and Welch, S. (Eds.) *Computational Methodologies for Gene Regulatory Networks*. IGI Publishers, New York, November 2009.

## **Refereed Book Chapters**

1. **Caragea, D.** and Xinming, Ou (2017). *Big Data Analytics for Mobile App Security*. In: Big Data Analytics in Cybersecurity (Data Analytics Applications), Onur Savas and Julia Deng (Eds.)

2. Herndon, N. and **Caragea, D.** (2015). *Empirical Study of Domain Adaptation Algorithms on the Task of Splice Site Prediction*. Invited paper. In: Lecture Notes in Communications in Computer and Information Science (CCIS), Springer-Verlag Berlin Heidelberg, vol. 511, pp. 195-211.
3. Herndon, N. and **Caragea, D.** (2014). *Predicting Protein Localization Using a Domain Adaptation Approach*. In Lecture Notes in Communications in Computer and Information Science (CCIS) 452, Springer-Verlag Berlin Heidelberg, pp. 191-206.
4. **Caragea, D.** and Honavar, V. (2009). *Learning Classifiers from Distributed Data Sources*. In: Encyclopedia of Database Technologies and Applications, Second Edition, Ferragline, V.E., Doorn, J.H., and Rivero, L.C. (Eds.), IGI Publishers, pp. 589-596.
5. Honavar, V. and **Caragea, D.** (2009). Invited Chapter. *Towards Semantics-Enabled Infrastructure for Knowledge Acquisition from Distributed Data*. In: Next Generation of Data Mining. Eds.: Kargupta, H., Han, J., Yu, P., Motwani, R., and Kumar, V. CRC Press. Ch. 16, pp. 317-337.
6. Paradesi, M.S.R., **Caragea, D.**, and Hsu, W.H. (2009). *Incorporating Graph Features for Predicting Protein-Protein Interactions*. In: Biological Data Mining in Protein Interaction Networks. X.-L. Li and S.-K. Ng (Eds.), IGI Publishers, pp. 45-63.
7. **Caragea, D.** and Honavar, V. (2009). *Knowledge Acquisition from Semantically Heterogeneous Data*. In: Encyclopedia of Data Warehousing and Mining, 2nd Edition, Wang, J. (Ed.). IGI Publishers, pp. 1110-1116.
8. **Caragea, D.**, Cook, D., Wickham, H., and Honavar, V. (2009). Invited Chapter. *Visual Methods for Examining SVM Classifiers*. In: Visual Data Mining: Theory, Techniques, and Tools for Visual Analytics. Springer, LNCS Volume 4404, pp. 136-153.
9. **Caragea, D.**, Silvescu, A., and Honavar, V. (2001). *Towards a Theoretical Framework for Analysis and Synthesis of Agents That Learn from Distributed Dynamic Data Sources*. Invited Chapter. In: Emerging Neural Architectures Based on Neuroscience. Pp. 547-559. Berlin: Springer-Verlag.

### Refereed Conference and Symposium Papers

1. Mazloom, R., Li, H., **Caragea, D.**, Imran, M., and Caragea, C. (2018). *Classification of Twitter Disaster Data Using a Hybrid Feature-Instance Adaptation Approach*. In: Proc. of the 15th Int. Conference on Information Systems for Crisis Response and Management (ISCRAM 2018), Rochester, NY, 2018. [WiPe paper. Acceptance rate N/A]
2. Neppalli, V.K., Caragea, C., and **Caragea, D.** (2018). *Deep Neural Networks versus Naive Bayes Classifiers for Identifying Informative Tweets during Disasters*. In: Proc. of the 15th Int. Conference on Information Systems for Crisis Response and Management (ISCRAM 2018), Rochester, NY, 2018. [WiPe paper. Acceptance rate N/A]
3. Sopova, O., Li, H., and **Caragea, D.** (2017). *Comparison of Two Domain Adaptation Approaches for Classifying Disaster-related Twitter Data*. In: Proceedings of the 4th International Symposium on Social Networks Analysis, Management and Security (SNAMS 2017), part of IEEE International Conference on Future Internet of Things and Cloud (FiCloud-2017), Prague, Czech Republic, 2017. [Acceptance rate: 30%]
4. Li, H., **Caragea, D.** and Caragea, C. (2017). *Towards Practical Usage of Domain Adaptation Algorithms in Classifying Disaster Related Tweets*. In: Proc. of the 14th Int. Conference on Information Systems for Crisis Response and Management (ISCRAM 2017), Albi, France, 2017. [WiPe paper. Acceptance rate 60%]

5. Stanescu, A., Tangirala, K. and **Caragea, D.** (2016). *Study of transductive learning and unsupervised feature construction methods for biological sequence classification*. **Best Paper Award**. In: Proceedings of the International Symposium on Network Enabled Health Informatics, Biomedicine and Bioinformatics, part of the Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2016) San Francisco, CA [Regular paper. Acceptance rate 30%]
6. Neppalli, K., Cerqueira Medeiros, M., Caragea, C., **Caragea, D.**, Tapia, A., Halse, S. (2016). *Retweetability Analysis and Prediction during Hurricane Sandy*. In: Proceedings of the 13th International Conference on Information Systems for Crisis Response and Management (ISCRAM 2016), Brazil. [Regular paper. Acceptance rate: 41%]
7. Herndon, N. and **Caragea, D.**, (2016). *Ab initio Splice Site Prediction with Simple Domain Adaptation Classifiers*. In: Proceedings of the 7th IEEE International Conference on Bioinformatics Models, Methods and Algorithms (BIOINFORMATICS 2016), Rome, Italy. [Short paper. Acceptance rate: N/A]
8. Roy, S., DeLoach, J., Li, Y., Herndon, N., **Caragea, D.**, Ou, X., Ranganath, V. P., Li, H., and Guevara, N. (2015). *Experimental Study with Real-world Data for Android App Security Analysis using Machine Learning*. In: Proceedings of the 2015 Annual Computer Security Applications Conference (ACSAC 2015), pp. 81-90, Los Angeles, CA [Regular paper. Acceptance rate: 25%].
9. Stanescu, A., and **Caragea, D.** (2015). *Predicting Cassette Exons Using Transductive Learning Approaches*. In: Proceedings of the Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB 2015), pp. 495-502, Niagara Falls, Canada [Regular paper. Acceptance rate: N/A].
10. Herndon, N., and **Caragea, D.** (2015). *An Evaluation of Self-training Styles for Domain Adaptation on the Task of Splice Site Prediction*. **Best Paper Award**. In: Proceedings of the International Symposium on Network Enabled Health Informatics, Biomedicine and Bioinformatics (HI-BI-BI 2015), part of the Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2015), pp. 1042-1047, Paris, France [Regular paper. Acceptance rate: 35%].
11. Parimi, R. and **Caragea, D.** (2015). *How to Choose a Recommender System: Insights and Experiences for Large-scale User Personalization*. In: Proceedings of the 4th IEEE International Congress on Big Data, Application Track. New York, USA [Regular paper. Acceptance rate: N/A].
12. Parimi, R. and **Caragea, D.** (2015). *Cross-Domain Matrix Factorization for Multiple Implicit-Feedback Domains*. In: Proceedings of the International Workshop on Machine learning, Optimization and big Data (MOD 2015), pp. 80-92, Taormina - Sicily, Italy [Regular paper. Acceptance rate: 43%].
13. Herndon, N. and **Caragea, D.** (2015). *Domain Adaptation with Logistic Regression for Splice Site Prediction*. In: Proceedings of the 11th International Symposium on Bioinformatics Research and Application (ISBRA 2015), Norfolk, Virginia [Regular paper. Acceptance rate: 35%].
14. Tangirala, K., Herndon, N. and **Caragea, D.** (2015). *Community Detection-based Feature Construction for Protein Sequence Classification*. In: Proc. of the 11th International Symposium on Bioinformatics Research and Application (ISBRA 2015), Norfolk, Virginia [Regular paper. Acceptance rate: 35%].
15. Li, H., Guevara, N., Herndon, N., **Caragea, D.**, Neppalli, K., Caragea, C., Squicciarini, A. and Tapia, A. (2015). *Twitter Mining for Disaster Response: A Domain Adaptation Approach*. In: 12th

International Conference on Information Systems for Crisis Response and Management (ISCRAM 2015), Norway [Short paper. Acceptance rate: 70%].

16. Stanescu, A., and **Caragea, D.** (2014). *Ensemble-based semi-supervised learning approaches for imbalanced splice site datasets*. In: Proceedings of the International Conference on Bioinformatics and Biomedicine (BIBM 2014), Belfast, UK, pp. 432-437 [Regular paper. Acceptance rate: 19%].
17. Herndon, N., Tangirala, K., and **Caragea, D.** (2014). *Predicting Protein Localization Using a Domain Adaptation Naive Bayes Classifier with Burrows Wheeler Transform Features*. In: Proceedings of the International Conference on Bioinformatics and Biomedicine (BIBM 2014), pp. 501-504, Belfast, UK [Short paper. Acceptance rate: 38%].
18. Tangirala, K. and **Caragea, D.** (2014). *Community Detection-based Features for Sequence Classification*. In: Proceedings of the 5th ACM Conference on Bioinformatics and Computational Biology (ACM-BCB 2014), pp. 559-568, Newport Beach, CA. [Short paper. Acceptance rate: 50%].
19. Herndon, N., and **Caragea, D.** (2014). *Empirical study of domain adaptation with naive Bayes on the task of splice site prediction*. In: Proceedings of the 5th International Conference on Bioinformatics Models, Methods and Algorithms (BIOINFORMATICS 2014), pp. 57-67, Loire Valley, France. [Regular paper. Acceptance rate: 14%]. **Nominated for best paper award.**
20. Tangirala, K., and **Caragea, D.** (2014). *Generating Features Using Burrows Wheeler Transformation for Biological Sequences*. In: Proceedings of the 5th International Conference on Bioinformatics Models, Methods and Algorithms (BIOINFORMATICS 2014), pp. 196-203, Loire Valley, France. [Short paper. Acceptance rate: 34%].
21. Tangirala, K., and **Caragea, D.** (2014). *Semi-supervised Classification of Protein Sequences Using Burrows Wheeler Transformation-based Features*. In: Proceedings of the 5th International Conference on Bioinformatics and Computational Biology (BICoB 2014 ), Las Vegas, Nevada [Acceptance rate: N/A].
22. Stanescu, A., and **Caragea, D.** (2014). *Semi-supervised Self-training Approaches for Imbalanced Splice Site Datasets*. In: Proceedings of the 5th International Conference on Bioinformatics and Computational Biology (BICoB 2014 ), Las Vegas, Nevada [Acceptance rate: N/A].
23. Zomlot, L., Chandran, S., **Caragea, D.**, and Ou, X. (2013). *Aiding Intrusion Analysis Using Machine Learning*. In: Proceedings of the 12th International Conference on Machine Learning and Applications (ICMLA 2013), Special Session on Machine Learning Challenges in Cyber Security Applications, Miami, FL. [Acceptance rate for conference: 26%; for special session N/A].
24. Stanescu, A., Nagar, S. and **Caragea, D.** (2013). *A Hybrid Recommender System: User Profiling from Keywords and Ratings*. In: Proceedings of the 2013 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2013), Atlanta, GA [Regular paper].
25. Parimi, R., **Caragea,** and Wunderlich, D. (2013). *Economic Development through Business Profiling: A Text Analysis based Approach*. In: Proceedings of the 2013 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2013), Atlanta, GA [Short paper].
26. Cheeti, S., Stanescu, A. and **Caragea, D.** (2013). *Cross-Domain Sentiment Classification Using an Adapted Naive Bayes Approach and Features Derived From Syntax Trees*. In: Proceeding of the 5th International Conference on Knowledge Discovery and Information Retrieval (KDIR 2013), Vilamoura, Algarve, Portugal [Short paper, Acceptance rate: 35%].
27. Tangirala, K. and **Caragea, D.** (2013). *Extraction of Gene Regulatory Networks from Biological Literature*. In: Proceeding of the IEEE 3rd International Conference on Computational Advances in Bio and medical Sciences (ICCABS 2013), New Orleans, LA [Acceptance rate: 43%].

28. Parimi, R. and **Caragea, D.** (2013). *Pre-release Box-Office Success Prediction for Motion Pictures*. In: Proceeding of the 10th International Conference on Machine Learning and Data Mining (MLDM 2013), p. 571-585, New York, NY [Acceptance rate: 33%]. **Nominated for best paper award.**
29. Herndon, N., and **Caragea, D.** (2013). *Naive Bayes Domain Adaptation for Biological Sequences*. In: Proceedings of the 4th International Conference on Bioinformatics Models, Methods and Algorithms (BIOINFORMATICS 2013), Barcelona, Spain [Regular paper. Acceptance rate: 10%]. **Nominated for best student paper award.**
30. Stanescu, A., and **Caragea, D.** (2012). *Semi-Supervised Learning of Alternatively Spliced Exons Using Expectation Maximization Type Approaches*. In: Proceedings of the 3rd International Conference on Bioinformatics Models, Methods and Algorithms (BIOINFORMATICS 2012), Algarve, Portugal [Short paper. Acceptance rate: 37%].
31. Tangirala, K. and **Caragea, D.** (2011) *Semi-Supervised Learning of Alternative Splicing Events Using Co-Training*. In Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM'11), Atlanta, GA [Short paper. Acceptance rate: 39.13%].
32. Caragea, C., Silvescu, A., Kataria, S., **Caragea, D.** and Mitra, P. (2011). *Classifying Scientific Publications Using Abstract Features*. In Proceedings of the Ninth Symposium on Abstraction, Reformulation and Approximation (SARA 2011), Parador de Cardona, Catalonia, Spain.
33. Zhang, S., **Caragea, D.** and Ou, X. (2011). *An Empirical Study of Using the National Vulnerability Database to Predict Software Vulnerabilities*. In Proceedings of the 22nd International Conference on Database and Expert Systems Applications (DEXA 2011), Toulouse, France [Regular paper. Acceptance rate: 25%].
34. Parimi, R. and **Caragea, D.** (2011). *Predicting Friendship Links in Social Networks Using a Topic Modeling Approach*. In: Proceedings of the 15th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2011), Shenzhen, China [Regular paper. Acceptance rate: 9.7%].
35. Caragea, C., Silvescu, A., **Caragea, D.** and Honavar, V. (2010). *Abstraction-Augmented Markov Models*. In: Proceedings of the IEEE Conference on Data Mining (ICDM 2010). Sydney, Australia. [Regular paper. Acceptance rate: 9%].
36. Caragea, C., Silvescu, A., **Caragea, D.**, and Honavar, V. (2010). *Semi-Supervised Sequence Classification Using Abstraction Augmented Markov Models*. In: Proceedings of the ACM Conference on Bioinformatics and Computational Biology (ACM-BCB 2010). Niagara Falls, NY. [Regular paper. Acceptance rate: 28%]
37. Volkova, S., **Caragea, D.**, Hsu, W.H., Drouhard, J. and Fowles, L. (2010). *Boosting Biomedical Entity Extraction by using Syntactic Patterns for Semantic Relation Discovery*. In: Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence (WI'10), Toronto, Canada. [Regular paper. Acceptance rate: 16.6%]
38. Xia, J., **Caragea, D.** and Hsu, W. (2009). *Bi-Relational Network Analysis Using a Fast Random Walk with Restart*. In: Proceedings of the IEEE International Conference on Data Mining (ICDM 2009), Miami, FL. [Short paper. Acceptance rate: 18%]
39. Caragea, C., **Caragea, D.** and Honavar, V. (2009). *Learning Link-Based Classifiers from Ontology-Extended Textual Data*. In: Proceedings of the 21st International Conference on Tools with Artificial Intelligence (ICTAI 2009), New Jersey. [Regular paper]
40. Haridas, M. and **Caragea, D.** (2009) *Exploring Wikipedia and DMoz as Knowledge Bases for Engineering a User Interests Hierarchy for Social Network Applications*. In: Proceedings of the 8th

International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE 2009), Algarve, Portugal. [Short paper. Acceptance rate: 38%]

41. Caragea, C., **Caragea, D.** and Honavar, V. (2009). *Learning Link-Based Naive Bayes Classifiers from Ontology-Extended Distributed Data*. In: Proceedings of the 8th International Conference on Ontologies, DataBases, and Applications of Semantics (ODBASE 2009), Algarve, Portugal. [Short paper. Acceptance rate: 38%]
42. Kulkarni, S. and **Caragea, D.** (2009). *Towards Bridging the Web and the Semantic Web*. In: Proceedings of the 2009 IEEE/WIC/ACM International Conference on Web Intelligence (WI'09), Milan, Italy. [Regular paper. Acceptance rate: 16%]
43. Kulkarni, S. and **Caragea, D.** (2009). *Computation of the Semantic Relatedness between Words Using Concept Clouds*. In: Proceedings of the International Conference on Knowledge Discovery and Information Retrieval (KDIR), part of the International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K). Madeira, Portugal. [Short paper. Acceptance rate: 34%]
44. **Caragea, D.**, Bahirwani V., Aljandal, W. and Hsu, W. (2009). *Ontology-Based Link Prediction in the LiveJournal Social Network*. In: Proceedings of the Eighth Symposium on Abstraction, Reformulation and Approximation (SARA'09), Lake Arrowhead, CA.
45. Aljandal, W., Bahirwani, V., **Caragea, D.** and Hsu, H.W. (2009). *Ontology-Aware Classification and Association Rule Mining for Interest and Link Prediction in Social Networks*. In: Proceedings of the AAAI 2009 Spring Symposium on Social Semantic Web: Where Web 2.0 Meets Web 3.0, Stanford, CA.
46. Xia, J., **Caragea, D.** and Brown, S.J. (2008). *Exploring Alternative Splicing Features Using Support Vector Machines*. In: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine (BIBM'08), Philadelphia, PA. [Acceptance rate: 24%]
47. Koul, N., Bahirwani, V., Caragea, C., **Caragea, D.**, and Honavar, V. (2008). *Learning from Large Autonomous Data Sources Using Sufficient Statistics*. Short Paper. In: Proceedings of the International Conference on Web Intelligence (WI 2008), Sydney, Australia. [Short Paper. Acceptance rate: 39%]
48. Harmon, S., DeLoach, S., Robby, **Caragea, D.** (2008). *Leveraging Organizational Guidance Policies with Learning to Self-Tune Multiagent Systems*. In: Proceedings of the Second IEEE International Conference on Self-Adaption and Self-Organization (SASO'08). Venice, Italy. [Acceptance rate: 27%]
49. Aljandal, W.A., Bahirwani, V., **Caragea, D.**, Hsu, W.H. and Weninger, T. (2008). *Validation-based Normalization and Selection of Interestingness Measures for Association Rules*. In: ANNIE 2008.
50. Paradesi, M.S.R., **Caragea, D.**, and Hsu, W.H. (2007). *Structural Prediction of Protein-Protein Interactions in Saccharomyces cerevisiae*. In: Proceedings of the 2007 IEEE 7th International Symposium on Bioinformatics and BioEngineering (BIBE'07). Boston, MA.
51. Bao, J., **Caragea, D.**, and Honavar, V. (2006). *A Tableau-based Federated Reasoning Algorithm for Modular Ontologies*. In: Proceedings of the 2006 IEEE / WIC / ACM International Conference on Web Intelligence. pp. 404-410. Hong Kong. [Acceptance rate: 18%]
52. Bao, J., **Caragea, D.**, and Honavar, V. (2006). *On the Semantics of Linking and Importing in Modular Ontologies*. In: International Semantic Web Conference (ISWC 2006). Athens, Georgia, USA. [Regular paper, Acceptance rate: 23%]



53. **Caragea, D.**, Zhang, J., Pathak, J., and Honavar, V. (2006). *Learning Classifiers from Distributed, Ontology-Extended Data Sources*. In: Proceedings of the 8th International Conference on Data Warehousing and Knowledge Discovery (DaWaK 2006). Krakov, Poland. LNCS. Berlin: Springer.
54. Bao, J., **Caragea, D.**, and Honavar, V. (2006). *Modular Ontologies - A Formal Investigation of Semantics and Expressivity*. **Best paper award**. In: Proceedings of the First Asian Semantic Web Conference (ASWC 2006). Beijing, China. R. Mizoguchi, Z. Shi, and F. Giunchiglia (Eds.) LNCS 4185, pp. 616-631, Springer-Verlag. [Regular paper, Acceptance rate: 18%]
55. Wickham, H., **Caragea, D.** and Cook, D. (2006). *Exploring High-Dimensional Classification Boundaries*. In: Proceedings of the 38th Symposium on the Interface of Statistics, Computing Science, and Applications - Interface 2006: Massive Data Sets and Streams. May 24-27, 2006, Pasadena, CA, USA.
56. Bao, J., **Caragea, D.**, and Honavar, V. (2006). *Towards Collaborative Environments for Ontology Construction and Sharing*. In: Proceedings of the 2006 International Symposium on Collaborative Technologies and Systems (CTS 2006). May 14-17, 2006 Las Vegas, Nevada, USA.
57. **Caragea, D.**, Zhang, J., Bao, J., Pathak, J., and Honavar, V. (2005). Invited paper. *Algorithms and Software for Collaborative Discovery from Autonomous, Semantically Heterogeneous Information Sources* (Invited paper). In: Proceedings of the 16th International Conference on Algorithmic Learning Theory. Lecture Notes in Computer Science. Singapore. Vol. 3734, Pp. 13-44. Berlin: Springer-Verlag.
58. Zhang, J., **Caragea, D.** and Honavar, V. (2005). *Learning Ontology-Aware Classifiers*. In: Proceedings of the Eight International Conference on Discovery Science (DS 2005), Springer-Verlag Lecture Notes in Computer Science. October 8-11, 2005, Singapore. Vol. 3735, Pp. 308-321. Berlin: Springer-Verlag.
59. **Caragea, D.**, Pathak, J., and Honavar, V. (2004). *Learning Classifiers from Semantically Heterogeneous Data*. In: Proceedings of the Third International Conference on Ontologies, DataBases and Applications of Semantics for Large Scale Information Systems (ODBASE 2004), Springer-Verlag Lecture Notes in Computer Science. October 25-29, 2004, Agia Napa, Cyprus. Vol. 3291, Pp. 963-980. Springer-Verlag.
60. Cook, D., **Caragea, D.**, and Honavar, V. (2004). *Visualization for Classification Problems, with Examples Using Support Vector Machines*. In: Proceedings of Computational Statistics (COMPSTAT 2004), 16th Symposium of IASC, August 23-27, 2004, Prague, Czech Republic. Pp. 799-806. Springer-Verlag.
61. **Caragea, D.**, Cook, D. and Honavar, V. (2003). *Towards Simple, Easy-to-Understand, but Accurate Classifiers*. In: Proceedings of the Third IEEE International Conference on Data Mining (ICDM 2003), November 19-22, 2003, Melbourne, FL, USA. Pp. 497-500. IEEE Press.
62. Reinoso J., Silvescu, A., **Caragea, D.**, Pathak, J., and Honavar, V. (2003). *A Federated Query-Centric Approach to Information Extraction and Integration from Heterogeneous, Distributed and Autonomous Data Sources*. In: Proceedings of the 2003 IEEE International Conference on Information Reuse and Integration (IRI 2003), October 27-29, 2003, Las Vegas, NV, USA. Pp. 183-191. IEEE Press.
63. **Caragea, D.**, Silvescu, A., and Honavar, V. (2003). *Decision Tree Induction from Distributed Data Sources*. In: Proceedings of the Conference on Intelligent Systems Design and Applications (ICDA 2003), August 10-13, 2004, Tulsa, OK, USA. Pp. 341-350. Springer-Verlag.

64. **Caragea, D.**, Cook, D., and Honavar, V. (2001). *Gaining Insights into Support Vector Machine Classifiers Using Projection-Based Tour Methods*. In: Proceedings of the Conference on Knowledge Discovery and Data Mining (KDD 2001), August 26-29, San Francisco, CA, USA. Pp. 251-256. ACM Press.
65. Agapie, A. and **Caragea, D.** (1997). *Genetic Algorithms, Schemata Construction and Statistics*. In: Proceedings of the International Conference on Computational Intelligence, Theory and Applications, 5th Fuzzy Days, Dortmund, Germany. Pp. 16-23.

### Refereed Workshop Papers

1. DeLoach, J., and **Caragea, D.** *Twitter-Enhanced Android Malware Detection*. In: Proceedings of the 2017 International Workshop on Big Data Analytics for Cyber Intelligence and Defense (BDA4CID 2017), in conjunction with the 2017 IEEE International Conference on Big Data (IEEE BigData 2017), Boston, MA, USA, 2017. [Regular paper. Acceptance rate: N/A]
2. DeLoach, J., **Caragea, D.**, Ou, Xinming (2016). *Android Malware Detection with Weak Ground Truth Data*. In: Proceedings of the 3rd International Workshop on Pattern Mining and Application of Big Data (BigPMA). In conjunction with the 2016 IEEE International Conference on Big Data (IEEE BigData 2016). Washington D.C., USA, 2016. [Regular paper. Acceptance rate: N/A]
3. Li, Y., Sundaramurthy, S.C., Bardas, A.G., Ou, X., **Caragea, D.**, Hu, X. and Jang, J. (2015). *Experimental Study of Fuzzy Hashing in Malware Clustering Analysis*. In: Proceedings of the 8th Workshop on Cyber Security Experimentation and Test (CSET-15), Usenix, Washington D.C.
4. Parimi, R., and **Caragea, D.** (2015). *Leveraging Multiple Networks for Author Personalization*. In: Proceedings of the AAAI-15 Workshop on Scholarly Big Data: AI Perspectives, Challenges, and Ideas (AAAI-15 SBD), Austin, Texas [Acceptance rate: 21%].
5. Parimi, R., and **Caragea, D.** (2014). *Community Detection on Large Graph Datasets for Recommender Systems*. In: Proceedings of the ICDM-2014 Workshop on Data Mining in Networks (DaMNet 2014), Shenzhen, China. [Acceptance rate: 40%].
6. Volkova, S., **Caragea, D.**, Hsu, W., Bujuru, S. (2010). *Animal Disease Event Recognition and Classification*. In: Proceedings of the First International Workshop on Web Science and Information Exchange in the Medical Web (MedEx'10). Collocated with the World Wide Web Conference WWW-2010, Raleigh, NC.
7. Caragea, C., Silvescu, A., **Caragea, D.** and Honavar, V. (2009). *Abstraction Augmented Markov Models*. In: NIPS Workshop on Machine Learning in Computational Biology (MLCB). [Acceptance rate: 30%]
8. Bahirwani, V., **Caragea, D.**, Aljandal, W.A. and Hsu, W.H. (2008). *Ontology Engineering and Feature Construction for Predicting Friendship Links in the Live Journal Social Network*. In: Proceedings of the KDD 2008 Second Workshop on Social Network Mining and Analysis (SNA-KDD). Las Vegas, NV, August 2008. ACM Digital Library. [Regular paper. Acceptance rate: 35%]
9. Bao, J., **Caragea, D.** and Honavar, V. (2007). *Query Translation for Ontology-Extended Data Sources*. In: Proceedings of the AAAI Workshop on Semantic e-Science, Vancouver, Canada. [Acceptance rate: 33%]
10. **Caragea, D.**, Bao, J. and Honavar, V. (2007). *Learning Relational Bayesian Classifiers on the Semantic Web*. In: Proceedings of the IJCAI 2007 Workshop on Semantic Web for Collaborative Knowledge Acquisition (SWeCKa 2007). In conjunction with the Twentieth International Joint Conference on Artificial Intelligence, Hyderabad, India, January 2007. [Acceptance rate: 20%].

11. **Caragea, D.**, Bao, J. and Honavar, V. (2006). *A General Strategy for Knowledge Acquisition from Semantically Heterogeneous Data Sources*. In: Proceedings of the AAAI 2006 Fall Symposia. Arlington, VA.
12. Bao, J., Hu, Z., **Caragea, D.**, Reecy, J., and Honavar, V. (2006). *A Tool for Collaborative Construction of Large Biological Ontologies*. In: Fourth International Workshop on Biological Data Management (BIDM 2006). Krakov, Poland. IEEE Press.
13. Bao, J., **Caragea, D.**, and Honavar, V. (2006). *A Distributed Tableau Algorithm for Package-based Description Logics*. In: Proceedings of the Second International Workshop on Context Representation and Reasoning (CRR 2006). Riva del Garda, Italy.
14. Pathak, J., Koul, N., **Caragea, D.** and Honavar, V. (2005). *A Framework for Semantic Web Services Discovery*. In: Proceedings of the 7th ACM International Workshop on Web Information and Data Management (WIDM-2005), Bremen, Germany. Pp. 45-50. ACM press.
15. **Caragea, D.**, Bao, J., Pathak, J., Silvescu, A., Andorf., C., Dobbs, D., and Honavar, V. (2005). *Information Integration from Autonomous Biological Data Sources*. In: Proceedings of the 3rd International Workshop on Biological Data Management (BIDM 2005), DEXA Workshops 2005, Copenhagen, Denmark. Pp. 580-584. IEEE Computer Society.
16. **Caragea, D.**, Pathak, J., Bao, J., Silvescu, A., Andorf., C., Dobbs, D. and Honavar, V. (2005). *Information Integration and Knowledge Acquisition from Autonomous Biological Data Sources*. In: Proceedings of the 2nd International Workshop on Data Integration in Life Sciences (DILS 2005), San Diego, CA. Vol. 3615, pp. 175-190. Berlin: Springer-Verlag.
17. Pathak, J., **Caragea, D.** and Honavar, V. (2004). *Ontology Extended Component-Based Workflows: A Framework for Constructing Complex Workflows from Semantically Heterogeneous Software Components*. In: Proceedings of the VLDB-04 Second International Workshop on Semantic Web and Databases (SWDB 2004), Toronto, Canada. Vol. 3372, pp. 41-56. Springer-Verlag.
18. **Caragea, D.**, Reinoso, J., Silvescu, A. and Honavar, V. (2003). *Statistics Gathering for Learning from Distributed, Heterogeneous and Autonomous Data Sources*. In: Proceedings of the IJCAI International Workshop on Information Integration on the Web (IIWeb 2003). Acapulco, Mexico. Pp. 99-104.
19. **Caragea, D.**, Silvescu, A., and Honavar, V. (2000). *Agents that Learn from Distributed Dynamic Data Sources*. In: Proceedings of the Workshop on Learning Agents, Agents 2000/ECML 2000. Stone, P. and Sen, S. (Eds.) ECML. June 3, Barcelona, Spain.
20. **Caragea, D.**, Silvescu, A., and Honavar, V. (2000). *Towards a Theoretical Framework for Analysis and Synthesis of Distributed and Incremental Learning Agents*. In: Proceedings of the Workshop on Distributed and Parallel Knowledge Discovery. ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2000). August 20, Boston, MA, U.S.A.

### Refereed Extended Abstracts in Conferences

1. DeLoach, J. and **Caragea, D.** (2017). *Using Twitter Data to Improve Android Malware Detection*. In: 2017 Annual Computer Security Applications Conference (ACSAC-2017). Poster program.
2. Li, Y., Ekambaram, R., DeLoach, J., Wei, F., Hall, L., Ou, X. and **Caragea, D.** (2017). *Zero-Day Android Malware Detection Using Machine Learning*. In: 2017 Annual Computer Security Applications Conference (ACSAC-2017). Poster program.

3. DeLoach, J., **Caragea, D.**, Ou, Xinming (2017). *Android Malware Detection with Weak Ground Truth Data*. In: Proceedings of the AAAI-17 Student Abstract and Poster Program, AAAI 2017 (Poster presentation.)
4. Bahirwani, V. and **Caragea, D.** (2011) *Study on Regulatory Motifs in Arabidopsis thaliana*. In: Proceedings of the 2nd ACM Conference on Bioinformatics and Computational Biology (ACM-BCB'11), Poster program.
5. Elshamy, W., **Caragea, D.** and Hsu, W.H. (2010). KSU KDD: Word Sense Induction by Clustering in Topic Space. In: Proceedings of the SemEval workshop. Poster program. Collocated with the 48th Annual Meeting of the Association for Computational Linguistic (ACL'10), Uppsala.
6. Bahirwani, V. and **Caragea, D.** (2009). *Exploring Transcription Patterns in Arabidopsis thaliana*. In: Proceedings of the 13th Annual International Conference on Research in Computational Molecular Biology (RECOMB'09), Poster Program. Tucson, AZ.
7. Paradesi, M.S.R., **Caragea, D.**, and Hsu, W.H. (2007). *Structural Prediction of Protein-Protein Interactions in Saccharomyces cerevisiae*. In: Proceedings of the Annual Meeting of the International Society for Computational Biology (ISMB 2007), Poster Program, Vienna, Austria.
8. Bao, J., **Caragea, D.**, and Honavar, V. (2006). *Package-based Description Logics - Preliminary Results*. In: International Semantic Web Conference - Doctoral Consortium (ISWC-DC 2006). Athens, Georgia, USA.
9. **Caragea, D.** and Honavar, V. (2006). *Knowledge Discovery from Disparate Earth Data Sources*. Second NASA Data Mining Workshop: Issues and Applications in Earth Sciences. Poster Session. Pasadena, CA.
10. **Caragea, D.**, Bao, J., Pathak, J. and Honavar, V. (2005) *Ontology-based Information Integration using INDUS System*. In: The Program of the Eight Annual Bio-Ontologies Meeting (Bio-Ont SIG 2005). Poster Session. Detroit, MI.
11. Pathak, J., Bao, J., **Caragea, D.**, Silvescu, A., Andorf., C., Yan, C., Dobbs, D. and Honavar, V. (2005). *INDUS: A System for Information Integration and Knowledge Acquisition from Autonomous, Distributed, and Semantically Heterogeneous Data Sources*. The Annual Meeting of the International Society for Computational Biology (ISMB 2005), Demo Program, Detroit, MI.
12. **Caragea, D.**, Silvescu, A., Pathak, J., Bao, J., Andorf., C., Yan, C., Dobbs, D. and Honavar, V. (2005). *Knowledge Acquisition from Autonomous, Distributed, Semantically Heterogeneous Data Sources*. The Annual Meeting of the Int. Society for Computational Biology (ISMB 2005), Poster Program, Detroit, MI.
13. Caragea, C., **Caragea, D.** and Honavar, V. (2005). *Learning Support Vector Machine Classifiers from Distributed Data Sources*. In: Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI 2005), Student Abstract and Poster Program, Pittsburgh, Pennsylvania. Pp. 1602-1603.
14. **Caragea, D.**, Syeda-Mahmood, T. (2004). *Semantic API Matching for Automatic Service Composition*. In: Proceedings of the 13th International World Wide Web conference on Alternate track papers and posters, Poster session (WWW 2004), May 17-22, 2004, New York, NY, USA. Pp. 436-437. ACM Press.
15. **Caragea, D.** (2002). *Learning in Open-Ended Dynamic Distributed Environments*. In: Proceedings of the 18th National Conference on Artificial Intelligence (AAAI 2002), Doctoral Consortium Program. Edmonton, Alberta, Canada. Pp. 980. AAAI Press.

16. **Caragea, D.**, Silvescu, A., and Honavar, V. (2000). *Incremental and Distributed Learning Using Support Vector Machines*. In: Proceedings of the 17th National Conference on Artificial Intelligence (AAAI 2000), Student Abstract and Poster Program. Austin, TX. Pp. 1067. AAAI Press.

### Tutorials

1. Honavar, V. and **Caragea, D.** (2006). *Semantic Web Technologies for Collaborative Knowledge Acquisition*. In conjunction with the 2006 1st International Conference on Digital Information Management (ICDIM), December 06-08, 2006, Christ College, Bangalore, India.
2. Honavar, V. and **Caragea, D.** (2006). *Collaborative Knowledge Acquisition from Semantically Disparate, Distributed Data Sources*. In conjunction with the 2006 International Symposium on Collaborative Technologies and Systems (CTS 2006). May 14-17, 2006 Las Vegas, Nevada, USA.

### Posters and Presentations

1. Roy, S., DeLoach, J., Li, Y., **Caragea, D.**, Ou, X., Herndon, N., Ranganath, V. Li, H., and Guevara, N. (2016). *Experimental Study with Real-world Data for Android App Security Analysis using Machine Learning*. Presented by Jordan DeLoach at the UT Dallas Undergraduate Research Expo in Computer Science, February 2016. **Won second place.**
2. DeLoach, J., Ou, X., and **Caragea, D.** (2015). *Scalable Security for Millions of Apps: Utilizing Tiered-Machine Learning to Scale to Market-wide Solutions*. Presented by Jordan DeLoach at Kansas State University Open House, April 2015.
3. Carolan, J.C., **Caragea, D.**, Reese, J.C., Reeck, G.R., Mutti, N.S., Tagu, D., Edwards, O.R. and Wilkinson, T.L. (2011) *An insight into the salivary secretome of the pea aphid *Acyrtosiphon pisum**. Sixth International Symposium on Molecular Insect Science, Amsterdam, The Netherlands.
4. Stanescu, A., **Caragea, D.** and Brown, S. (2011) *Semi-Supervised Learning Approaches for Predicting Alternatively Spliced Exons*. Presented at CRA-W Grad Cohort Workshop, Boston, MA.
5. Volkova, S., **Caragea, D.** and Hsu, W. (2010) *Automated Event Extraction and Named Entity Recognition in the Domain of Veterinary Medicine*. Poster presented by Svitlana Volkova at the 2010 Grace Hopper Celebration, September 28 - October 2, Atlanta, GA.
6. Volkova, S., Hsu, W. and **Caragea, D.** (2009) *Named Entity Annotation and Tagging in the Domain of Epizootics*. Poster presented by S. Volkova at the Women in Machine Learning Workshop co-located with NIPS'09 Conference, December 2009, Vancouver, CA.
7. Narro, M., Ram, S., **Caragea, D.**, Cushing, J. and Brown, S. (2009) *Computing Opportunities in Plant Sciences*. Poster at 2009 Grace Hopper Celebration, September 30 - October 3, Tucson, AZ.
8. Johnson, L., Surabhi, G.C., Kumar, S., **Caragea, D.**, Lu, N., Shah, J. (2009). *Chronic and transient effects of nitrogen saturation on root processes in a dominant prairie grass *Andropogon gerardii*: Linking gene expression profiles and ecological responses*. Poster presented by Johnson at the Plant and Animal Genomics meeting, January 2009, San Diego.
9. Bahirwani, V. and **Caragea, D.** (2009). *Exploring Transcription Patterns in *Arabidopsis thaliana**. In: Proceedings of the 13th Annual International Conference on Research in Computational Molecular Biology (RECOMB'09), Poster Program. Tucson, AZ.
10. Bahirwani, V., Tolos, S., **Caragea, D.**, Brown, S. and Wang, H. (2008). *Computational Methods to Characterize Regulatory Networks involved in Plant Response to Abiotic Stresses*. Poster presented at the 6th KSU Ecological Genomics Annual Symposium November 14 - 16, 2008 in Kansas City.

11. Wang, H., **Caragea, D.** and Susan J. Brown (2008). *Computational Methods to Characterize Regulatory Networks Involved in Plant Response to Abiotic Stresses*. Oral Presentation at KSU Ecological Genomics Research Forum, May 2008.
12. **Caragea, D.**, Kallumadi, S., Dittmer, N., Chellapilla, S., Mutti, N., Feng, C., Pierson, M., Heerman, M., Culbertson, C., Reese, J., Edwards, O. and Reeck, G. (2008). *Identifying Specialized Salivary Gland Transcripts in Pea Aphid Using Bioinformatics Tools*. Poster presented at the Second Annual Arthropod Genomics Symposium: New Insights from Arthropod Genomes, April 11 - 13, 2008, in Kansas City.
13. Chellapilla, S., Kallumadi, S., Park, Y., **Caragea, D.** and Brown, S.J. (2008). *ArthropodEST: A Pipeline for Automated EST Data Analysis*. Poster presented at the Second Annual Arthropod Genomics Symposium: New Insights from Arthropod Genomes, April 11 - 13, 2008, in Kansas City.
14. Cui, F., Dai, H., Hiromasa, Y., **Caragea, D.**, Sheng, C., Reese, J., Edwards, O. and Reeck, G. (2008). *Characterization of an endoplasmic reticulum protein from the salivary glands of the pea aphid, *Acyrtosiphon pisum**. Poster presented at the Second Annual Arthropod Genomics Symposium: New Insights from Arthropod Genomes, April 11 - 13, 2008, in Kansas City.
15. Steller, M., Kambhampati, S., and **Caragea, D.** (2008). *Bioinformatic Analysis of ESTs from Termite Castes*. Poster presented at the Second Annual Arthropod Genomics Symposium: New Insights from Arthropod Genomes, April 11 - 13, 2008, in Kansas City.
16. Surabhi, G.C., Kumar, S., Alam, N., **Caragea, D.**, Lu, N., Hurt, A., Johnson, L., Shah, J. (2008). *Chronic and transient effects of nitrogen saturation on root processes in a dominant prairie grass *Andropogon gerardii*: Linking gene expression profiles and ecological responses*. Poster presented at the Plant Biology Meeting, 2008, Mexico.
17. Paradesi, M.S.R., **Caragea, D.**, and Hsu, W.H. (2007). *Structural Prediction of Protein-Protein Interactions in *Saccharomyces cerevisiae**. In: Proceedings of the Annual Meeting of the International Society for Computational Biology (ISMB 2007), Poster Program, Vienna, Austria.
18. Paradesi, M., Hsu, W. and **Caragea, D.** (2006). *Protein-Protein Interaction Prediction*. Poster presented at the Kansas Bioscience Community meeting held Monday November 13th, 2006 at KSU.
19. Honavar, V. and **Caragea, D.** (2006). *Querying Semantically Heterogeneous Data Sources from a User's Point of View*. 2006 Semantic Technology Conference. San Jose, CA, March 6-9, 2006.
20. **Caragea, D.** and Honavar, V. (2004). *Knowledge Acquisition from Semantically Heterogeneous Distributed Data*. Demo Presentation at the NSF Information and Data Management Workshop (IDM), Cambridge, Massachusetts, October 10-12, 2004.
21. Jie, B., Yan, C., **Caragea, D.** and Honavar, V. (2004). *Integration of Ontology-Extended Biological Data Sources*. Poster presented at Standards and Ontologies for Functional Genomics, Philadelphia, PA, 2004.
22. **Caragea, D.**, Silvescu, A., and Honavar, V. (2000). *Multi-Agent Learning from Distributed Data Sources*. In: Workshop on Multi-Agent Learning: Theory and Practice, organized by G. Tesauro and A. Greenwald. International Conference on Machine Learning (ICML-2000), Stanford University.

#### **K-INBRE GRA Summit Presentations**

1. Tangirala, K. and **Caragea, D.** (2011) *Semi-supervised Learning of Alternative Splicing Events*. Fourth K-INBRE Bioinformatics GRA Summit, May 2011, Lawrence, KS.

2. Nagar, S., **Caragea, D.** and Brown, S.J. (2011) *AgriPest: a Gene Annotation Portal*, Fourth K-INBRE Bioinformatics GRA Summit, May 2011, Lawrence, KS.
3. Rao, H.S., Visvanathan, M., **Caragea, D.**, Bousfield, G.R., and Lushington, G.H. (2010). *Mining Microarray Data for Classification of Differentially Expressed Genes Implicated in Follicle Stimulating Hormone Specificity in Male Rat*. Third K-INBRE Bioinformatics GRA Summit, January 2010, Kansas City, KS.
4. Tangirala, K., **Caragea, D.**, Bahirwani, V., Brown, S., Welch, S. (2010) *Extracting Gene Regulatory Networks from Literature*. Third K-INBRE Bioinformatics GRA Summit, January 2010, Kansas City, KS.
5. Nagar, S., **Caragea, D.**, Brown, S. (2010) *Textpresso for Red Flour Beetle*. Third K-INBRE Bioinformatics GRA Summit, January 2010, Kansas City, KS.
6. Bahirwani, V., **Caragea, D.**, Brown, S. (2010) *Exploring Transcription Patterns and Regulatory Motifs in Arabidopsis thaliana*. Third K-INBRE Bioinformatics GRA Summit, January 2010, Kansas City, KS.
7. El-Sherif, **Caragea, D.**, Brown, S. (2010) *Modeling Segmentation Mechanisms in Tribolium castaneum*. Third K-INBRE Bioinformatics GRA Summit, January 2010, Kansas City, KS.
8. Herndon, N., Osei-Boadi, K., Jiang, D., **Caragea, D.**, Brown, S. (2010) *MAKER and Apollo: tools for genome annotation*. Second K-INBRE Bioinformatics GRA Summit, July 2010, Manhattan, KS.
9. Mungre, S., Tangirala, K., **Caragea, D.**, Park, Y. and Brown, S. (2010) *A Database for Tick Transcriptomics*. Second K-INBRE Bioinformatics GRA Summit, July 2010, Manhattan, KS.
10. Maurer, D., Zheng, L., **Caragea, D.**, Park, Y., and Brown, S. (2010) *Assessing genome wide expression using DNA tiling arrays*. Second K-INBRE Bioinformatics GRA Summit, July 2010, Manhattan, KS.
11. Nagar, S., Vaswani, V., **Caragea, D.**, Park, Y. and Brown, S. (2010) *Introduction to the CHADO schema for genomic databases and JBrowse the fast genome browser*. Second K-INBRE Bioinformatics GRA Summit, July 2010, Manhattan, KS.
12. Kulkarni, S., Bahirwani, V., **Caragea, D.**, Brown, S. (2009) *Mining interesting information from online data sources for Arabidopsis thaliana*. K-INBRE Bioinformatics GRA Summit, May 2009, Lawrence, KS.
13. Bahirwani, V., **Caragea, D.** and Brown, S. (2009) *From dependency networks for gene regulatory networks in Arabidopsis thaliana*. First K-INBRE Bioinformatics GRA Summit, May 2009, Lawrence, KS.

### Invited Seminars and Presentations

1. **Caragea, D.** (2017) *Leveraging Multiple Networks for Enhanced Recommendations*. Invited talk in the Mini-Symposium “MS15: Interactions Among Analysis, Optimization and Network Science” at the 3rd Annual Meeting of SIAM Central States Section, Colorado State University, September 29 - October 1, 2017.
2. **Caragea, D.** (2017) *Graph-based Recommender Systems*. Invited seminar for the NODE group, Math Department, Kansas State University, May 2017.

3. **Caragea, D.** (2017) *Mining Twitter to Aid Disaster Response*. Invited talk at the Kansas City Machine Learning Group monthly meeting, February 2017.
4. **Caragea, D.** (2016) *Learning Domain Adaptation Classifiers from Multiple Distributed Sources*. Invited talk in the session on “Inference and prediction for distributed data” at the 2016 Annual Meeting of the Statistical Society of Canada, May 2016.
5. **Caragea, D.** (2016) *Mining Twitter to Aid Disaster Response*. Invited talk at the Reunion and Seminar “Honoring Beth,” Kansas State University, April 2016.
6. **Caragea, D.** (2015) *Visual Methods for Examining Machine Learning Classifiers*. Invited speaker. First Midwest Bioinformatics Conference, Kansas City, October 15-16, 2015.
7. **Caragea, D.** (2015) *Semi-Supervised Learning and Domain Adaptation Approaches in Bioinformatics: Challenges and Opportunities*. Invited seminar. Department of Biochemistry and Molecular Biophysics, Kansas State University, February 18th, 2015.
8. **Caragea, D.** (2009) *Biological Sequence Classification Using Machine Learning*. Seminar in Computer Science Department at Kansas State University, Fall 2009.
9. **Caragea, D.** (2009) *Biological Sequence Classification Using Machine Learning*. Invited seminar. Plant Pathology, Kansas State University, Fall 2009.
10. **Caragea, D.** (2008). *Ontology-Based Link Prediction in the LiveJournal Social Network*. Talk at IBM Research, Watson, NY, October 24th, 2008.
11. **Caragea, D.** (2008). *Computational Methods for Gene Regulatory Network Discovery*. Invited seminar. L.H. Baker Center for Bioinformatics and Biological and Computational and Systems Biology Summer Institute, Iowa State University, June 14th, 2008.
12. **Caragea, D.** (2008). *Visual Methods for Examining Support Vector Machine Results*. Invited seminar. Department of Statistics, Kansas State University, April 4th 2008.
13. **Caragea, D.** (2008). *Knowledge Discovery from Biomedical Data Sources*. Cerner Corporation, March 11th, 2008.
14. Garrett, K., Tang, Z., **Caragea, D.** and Gadbury, G. (2008). *Statistical Workshop for Microarray Data Analysis*. Training Workshop. Organized by Karen Garrett at The Centennial Meeting of the American Phytopathological Society (APS), July, 2008, Minneapolis, MN.
15. **Caragea, D.** (2006). *Collaborative Knowledge Acquisition from Distributed, Semantically Heterogeneous Data Sources*, Technology Screening Panel, Davenport, IA, February 16, 2006. Invited Talk. New Ventures AgTech Initiative, Eastern Iowa Community College District.
16. **Caragea, D.** (2005). *Information Integration and Knowledge Acquisition from Semantically Heterogeneous Biological Data Sources*. Division of Biomedical Informatics at Cincinnati Children’s Hospital Medical Center, Cincinnati, OH, October 21st, 2005. Invited Talk.
17. **Caragea, D.** (2004). *Learning Classifiers from Distributed, Semantically Heterogeneous, Autonomous Data Sources*. Multimedia and Database Research Group, Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, November 5th 2004.
18. **Caragea, D.** (2004). *Learning Classifiers from Distributed, Semantically Heterogeneous, Autonomous Data Sources*. The IBM Data Analytics Research Group, T.J. Watson Research Center, Yorktown Heights, NY, October 4th, 2004.



19. **Caragea, D.** (2004). *Learning Classifiers from Distributed, Semantically Heterogeneous, Autonomous Data Sources*. Colloquium Series Spring 2004, Department of Computer Science, Iowa State University, Ames, IA, July 27th, 2004.

### **Professional Affiliations**

- Member, American Association for Artificial Intelligence (AAAI)
- Member, Association for Computing Machinery (ACM)
- Member, International Society for Computational Biology (ISCB)

### **Department & University Service**

- Member, CS Faculty Search Committee (2014-present)
- Member, College Honors Program Committee (2011-present)
- Chair, CS Ph.D. Fellowship Committee (2014-present)
- Faculty advisor, ACM-W local chapter (2014-present)
- Member, CS Ph.D. Fellowship Committee (2009-2014)
- Chair, KSU Bioinformatics Curriculum Committee (2006-2014)
- Member, CS Seminar Series Committee (2006-2014)
- Member, CS Department Website Committee (2007-2014)
- Member, Bioinformatics faculty-hiring committees, Plant Pathology, KSU, 2014.
- Member, Bioinformatics faculty-hiring committees, Entomology, KSU, 2014.
- Member, CS Business Manager Search Committee, May 2011
- Member, CS Eco-forecasting - Two Research Assistant Search Committees, Spring and Fall 2010
- Member, iPlant Postdoc Search Committee, KSU, Fall 2009
- Member, CS Chair Search Committee (2008-2009)
- Member, AGC Data Analyst Search Committee, Arthropod Genomics Center, January 2007
- Member, Bioinformatics Specialist Search Committee, Bioinformatics Center, June 2007

### **STEM Outreach**

- Helped recruit students for the “Missouri, Iowa, Nebraska, Kansas Women in Computing (MINK WIC)” conference, which was held in Kansas City, Missouri in 2011 and 2015. MINK WIC is being funded through a Grace Hopper Regional Consortium grant.
- Participated in the Developing Scholars program - mentored freshman Tiana Brooks from ECE on a research project on “Data Mining of Social Networks” (2010-2011); mentored freshman Geordy Williams from CS on a research project on “Using Recommender Systems for Social Network Applications” (2013-2014); mentored freshman Trevin Garcia from CS on a research project on “News Aggregation Approaches” (2014-2015).
- Co-organized two campus-wide genomics workshops: “Hessian Fly Genboree” (May 5, 2010) and “Annotate Your Favorite Gene” (June 8-9, 2010), together with Susan J. Brown in Biology.
- Offered a campus-wide two-day hands-on workshop, called “It’s a BLAST!” (May 2009), in collaboration with Susan J. Brown in Biology. The workshop was focused on pair-wise alignments, dynamic programming and database search algorithms
- Co-organized and coordinated a Bioinformatics Summer Internship Program, together with Susan J. Brown in Biology. Nine graduate students from three departments participated in this program.
- Honored with three *WESP Making a Difference Awards* (nominated by three female students for making a difference for them during their time at K-State) in December 2009 and May 2010.
- Attended an afternoon social event focused on Arthropod Genomics at the Manhattan/K-State Innovation Center (discussion led by the president Kirk Schulz), Fall 2010.
- Attended the KSU Energy Summit (meant to examine the potential for an interdisciplinary center at the nexus of energy, water, and land), Fall 2010.
- Participated in a Society of Women in Engineering (SWE) female professor panel, November 2010.
- Participated in the Engineering Scholarship Day Seminar, October 2009.
- Participated as a panelist in the ADVANCE Distinguished Lecture Series Luncheon, January 2009.
- Invited presenter at the second annual *STAR (Science, Technological Advancements and Research) Tech Conference*, held on Monday, January 19th, 2009 at Salina High School South, Salina, Kansas. Talk on “*Bioinformatics: How will it help society?*”
- Participated in a Workshop on “COACHing Strong Women in the Power of Strategic Persuasion” (Fall 2008).
- Presenter, KSU Girls Researching Our World workshop (2007, 2008). Lectures on “*What makes blood red?*”
- Group Leader, WESP Learning Community discussion group for graduate female students (book discussed: *How to complete and survive a Ph.D. Thesis*).

## Professional Activities

### Grant Reviewing

- Reviewer and panelist for NSF, CISE Directorate, IIS Division: 2008 (1 panel), 2009 (3 panels), 2010 (1 panel), 2011 (1 panel), 2016 (1 panel), 2017 (2 panels), 2018 (2 panels)

- Reviewer and panelist for the American Association for the Advancement of Science (AAAS), 2016.
- Reviewer for the Maryland Industrial Partnerships Program (MIPS), Maryland, 2007.
- Reviewer for the Air Force Office of Scientific Research, Chemistry and Life Sciences Program, 2007.
- Reviewer for the Science & Engineering Research Council (SERC), Singapore, 2005.

### **Organizing Committees**

- Co-chair, AAAI Second Workshop on Scholarly Big Data: AI Perspectives, Challenges, and Ideas (AAAI-SBD 2016). In conjunction with the Thirtieth AAAI Conference (AAAI 2016), Phoenix, AZ, February 2016.
- Organizing Committee Member, AAAI Workshop on Scholarly Big Data: AI Perspectives, Challenges, and Ideas (AAAI-SBD 2015). In conjunction with the Twenty Ninth AAAI Conference (AAAI 2015), Austin, TX, January 2015.
- Organizing Committee Member, Workshop on Semantic Web for Collaborative Knowledge Acquisition (SWeCKa 2007). In conjunction with the Twentieth International Joint Conference on Artificial Intelligence (IJCAI 2007), Hyderabad, India, January 2007.
- Organizing Committee Member, The 2006 AAAI Fall Symposium on Semantic Web for Collaborative Knowledge Acquisition (SWeCKa 2006), Arlington, VA, October 2006.
- Organizing Committee Chair, The IEEE Workshop on Knowledge Acquisition from Distributed, Autonomous, Semantically Heterogeneous Data and Knowledge Sources. In conjunction with the Fifth IEEE International Conference on Data Mining (ICDM 2005), Houston, TX, November 2005.

### **Conference Session Chair**

- Session Chair at the First ACM International Conference on Bioinformatics and Computational Biology (ACM-BCB), Niagara Falls, August 2010.
- Session Chair at the 4th Annual Arthropod Genomics Symposium Arthropod Genomics: New Approaches and Outcomes, Kansas City, June 2010.
- Session Chair at the 2009 International Conference on Knowledge Discovery and Information Retrieval (KDIR 2009), Madeira, Portugal, October 2009.
- Session Chair at the National Science Foundation Symposium on Next Generation of Data Mining and Cyber-Enabled Discovery for Innovation, October 2007, Baltimore, MD.

### **Program Committees and Conference Reviewing**

- Program Committee Member, International Conference on Social Networks and Social Media (SNSM'17), 2017.
- Program Committee Member, IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2007-2018.
- Program Committee Member, International Symposium on Network Enabled Health Informatics, Bio-Medicine and Bioinformatics (HI-BI-BI), 2013-2018.
- Program Committee Member, International Symposium on Bioinformatics Research and Applications (ISBRA), 2008, 2012-2018.

- Program Committee Member, International Conference on Computational Linguistics (COLING 2016)
- Program Committee Member, International Conference on Knowledge Engineering and Ontology Development (KEOD), 2011-2016.
- Program Committee Member, ACL 2015 Workshop on Novel Computational Approaches to Keyphrase Extraction (ACL-Keyphrase), 2015.
- Program Committee Member, International Symposium on Resilient Cyber Systems, 2015.
- Program Committee Member, First IEEE International Conference on Computational Advances in Bio and Medical Sciences (ICCABS), 2011-2014.
- Program Committee Member, Computational Scientometrics Workshop, 2013.
- Program Committee Member, Conference on Statistical, Computational, and Visualization Methods in Medical Informatics, 2011.
- Program Committee Member, First ACM International Conference on Bioinformatics and Computational Biology (ACM-BCB), 2010.
- Program Committee Member, Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), 2009-2010.
- Program Committee Member, International Joint Conferences on Bioinformatics, Systems Biology and Computational Intelligence (IJCBS), 2009.
- Program Committee Member, Tenth Argentine Symposium on Artificial Intelligence (ASAI), 2009.
- Program Committee Member, Midwest Artificial Intelligence and Cognitive Science Conference (MAICS), 2003-2007, 2009.
- Program Committee Member, AAAI Doctoral Consortium (DC) Program, in conjunction with The National Conference on Artificial Intelligence (AAAI), 2006-2008, 2010-2011.
- Program Committee Member, SIAM Conference on Data Mining (SDM), 2008.
- Program Committee Member, Workshop on Collaborative Distributed Knowledge Discovery (CDKD). In conjunction with the International Symposium on Collaborative Technologies and Systems (CTS), 2008.
- Program Committee Member, Workshop on Visual Data Mining (VDM). In conjunction with the Second International Conference MCO, 2008.
- Program Committee Member, International Conference on BioMedical Engineering and Informatics (BMEI), 2008.
- Program Committee Member and Special Area Chair (Bio-Ontologies), The International Conference on Biocomputation, Bioinformatics, and Biomedical Technologies (BIOTECHNO), 2008.
- Program Committee Member, International Conference on Advanced Engineering Computing and Applications in Sciences (ADVCOMP), 2008.
- Program Committee Member, International Conference on Autonomic and Autonomous Systems (ICAS), 2006-2008.

- Program Committee Member, International Conference on Computing in the Global Information Technology (ICCGI), 2006-2008.
- Program Committee Member, Workshop on Cyberinfrastructure for e-Science (CyIneS). In conjunction with the 2007 IEEE/WIC/ACM Conference on Web Intelligence (WI), 2007.
- Program Committee Member, International Conference on Advances in Semantic Processing (SEMAPRO), 2007.
- Program Committee Member, International Workshop on Healthcare Information and Knowledge Management (HIKM). In conjunction with the ACM Fifteen Conference on Information and Knowledge Management (CIKM), 2006.
- Program Committee Member, AAAI Student Abstracts and Poster Program (SA), in conjunction with the Twenty-First National Conference on Artificial Intelligence (AAAI), 2006.
- Program Committee Member, International Workshop on Contexts and Ontologies: Theory, Practice and Applications (C&O), in conj. with the 17th European Conference on Artificial Intelligence (ECAI), 2006.
- Program Committee Member, Workshop on Collaborative Knowledge Discovery in Databases (CKDD) in conjunction with the International Symposium on Collaborative Technologies and Systems (CTS), 2006.
- Technical Committee Member, International Conference on Intelligent Knowledge Systems, 2005.

### **Journal Article Reviewing**

- Information Processing Letters (Elsevier Journal), 2017.
- -2.5mm Information Sciences, 2017.
- SpringerPlus, 2016.
- Mathematical Problems in Engineering, 2016.
- Journal of Contingencies and Crisis Management (JCCM), Special Issue on HCI in Critical Systems, 2016.
- Computational Biology and Chemistry (Eviser), 2016.
- METHODS (Elsevier Journal), 2015.
- Neurocomputing (Elsevier Journal), 2015, 2016.
- Plos One, 2012, 2015, 2016.
- International Journal of Molecular Sciences (IJMS), 2015.
- IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB), 2008-2010, 2015.
- BMS Genetics, 2014.
- IEEE Transactions on Software Engineering (TSE), 2014.
- BMC Genomics, 2011, 2012, 2013.
- ISRN Computational Biology, 2013.

- Journal of Multimedia Tools and Applications, 2009, 2013.
- BMC Genomics, 2011.
- Wiley Interdisciplinary Reviews (WIREs): Data Mining and Knowledge Discovery (DMKD), 2007, 2011.
- International Journal on Data Mining and Bioinformatics (IJDMB), 2010.
- IEEE Transactions on Neural Networks (TNN), 2010.
- IEEE Transactions on Data and Knowledge Engineering (TKDE), 2007-2009.
- Elsevier Journal of Data & Knowledge Engineering (DKE), 2005, 2009.
- Expert Systems, 2007-2009.
- ACM Transactions on Knowledge Discovery from Data (TKDD), 2008.
- International Journal of Intelligent Information and Database Systems, 2008.
- Pattern Recognition Letters (PRL), 2006-2007.
- Journal of Research and Practice in Information Technology (JRPIT), 2007.
- World Enformatika Society (www.enformatika.org), 2006.
- International Journal of Computer Science (IJCS), 2006.

### **Book Chapter Reviewing**

- *Computational Methodologies in Gene Regulatory Networks*, Das, S., Caragea, D., Hsu, W.H., Welch, S. (Eds.). IGI Global, 2009. [Reviewed 9 chapters]
- *Encyclopedia of Database Technologies and Applications*, Second Edition, Ferraggine, V.E., Doorn, J.H., and Rivero, L.C. (Eds.). IGI Global, 2008. [Reviewed 2 chapters]
- *Encyclopedia of Data Warehousing and Mining*, Second Edition, Wang, J. (Ed.). IGI Global, 2008. [Reviewed 2 chapters]
- *Biological Data Mining in Protein Interaction Networks*, Li, X.L. and Ng, S.K. (Eds.). IGI Global, 2008. [Reviewed 1 chapter]
- *Visual Data Mining: Theory, Techniques, and Tools for Visual Analytics*. Simoff, Simeon; Bhlen, Michael; Mazeika, Arturas (Eds.) Springer, LNCS Volume 4404. [Reviewed 1 chapter]
- *Next Generation of Data-Mining (1)*, Eds. Mehmed Kantardzic and Jozef Zurada, Wiley-IEEE Press, 2005. [Reviewed 1 chapter]