# Cornelia Caragea

## Office address

226 Atanasoff Hall
Computer Science Department
Iowa State University
Ames, Iowa 50011-1041
Phone: (515) 294-4377

Phone(office): (515)294-9074
Phone(home): (515) 451-3997
Email: cornelia@cs.iastate.edu
URL: http://www.cs.iastate.edu/ cornelia

## Education

2003-Present: **Ph.D. in Computer Science, Iowa State University, USA**
Advisor: **Vasant Honavar**

1992-1997: **B.S. in Computer Science, University of Bucharest, Romania**
Thesis: *Operational Research, Constraints Optimization*
Advisor: **Vasile Preda**

## Computer skills

**Programming languages:**
Java, C, C++, Perl, Excel, Matlab (can become proficient in other languages)

**Operating Systems:**
Unix, Linux, Windows, Mac OS X, MS DOS

**Other**
LaTeX

## Research areas of interest

Machine Learning, Data Mining and Knowledge Discovery, Relational Learning, Probabilistic Graphical Models, Artificial Intelligence, Bioinformatics and Computational Biology

## Awards and Honors

**Leonardo Melandri Program Award** *December 10-15, 2006*
DASI '06: Bertinoro PhD School on Data and Service Integration, Bertinoro, Italy.
Presented "INDUS: Intelligent Data Understanding System", a system developed by our group.

**ISU Travel Award** *December 1-3, 2006*
Award given by the Computer Science Department and Iowa State University to attend the 4th Annual Rocky Mountain Bioinformatics Conference, Aspen/Snowmass, Colorado.

**AAAI Student Scholarship** *July 9-13, 2005*
Award given by the Twentieth National Conference on Artificial Intelligence, Pittsburgh, Pennsylvania.

**John Vincent Atanasoff Award** *2005*
Annual award given by the Computer Science Department to outstanding graduate students.

**PACE Award** *2003-2004*
Premium for Academic Excellence from Iowa State University.

## Professional Activities

**Reviewer**
Indian International Conference on Artificial Intelligence

**Professional Societies Membership**
Member of American Association for Artificial Intelligence (AAAI)

## Mentored Students

Carlos M Herrera Jr.

David Gemperline

## Publications and Research Reports

**Journal Papers:**

**Caragea, C.**, Sinapov, J., Silvescu, A., Dobbs, D., and Honavar, V. **Glycosylation site prediction using ensembles of Support Vector Machine classifiers**, *BMC Bioinformatics 2007*, 8:438

**Conference Papers:**

Lee. J-H., Hamilton, M., Gleeson, C., **Caragea, C.**, Zaback, P., Sander, J., Lee, X., Wu, F., Terribilini, M., Honavar, V. and Dobbs, D. **Striking Similarities in Diverse Telomerase Proteins Revealed by Combining Structure Prediction and Machine Learning Approaches**, In Proceedings of the *Pacific Symposium on Biocomputing* (PSB 2008)

**Caragea, C.**, Sinapov, J, Dobbs, D., and Honavar, V. **Assessing the Performance of Macromolecular Sequence Classifiers**, In Proceedings of the *IEEE Conference on Bioinformatics and Bioengineering* (BIBE 2007)

**Workshop Papers:**

Towfic, F., Gemperline, D.C., **Caragea, C.**, Wu, F., Dobbs, D., and Honavar, V. **Structural Characterization of RNA-Binding Sites of Proteins: Preliminary Results**, In *Computational Structural Bioinformatics Workshop, IEEE International Conference on Bioinformatics and Biomedicine* (BIBM 2007)

**Student Abstracts and Poster Programs:**

Reyon, D., Lewis, B., Lee, J.H., Gleeson, C., Hamilton, M., **Caragea, C.**, Towfic, F., Terribilini, M., Honavar, V., and Dobbs, D. **Combining Structural Modeling, Evolutionary Information, and Machine Learning to Improve Prediction of Nucleic Acid Binding Sites in Telomerase.** Accepted to The 5th Annual Rocky Mountain Bioinformatics Conference, Poster Program, Aspen/Snowmass, Colorado (Rocky 2007)

Terribilini, M., Sander, J., **Caragea, C.**, Lee, J-H., Jernigan, R.L., Honavar, V. and Dobbs, D. **Comparing Sequence vs. Structure-based Predictions of RNA Binding Sites in Proteins.** Midwest Symposium on Computational Biology and Bioinformatics, Poster Program, Chicago, Illinois (MSCBB 2007)

Towfic, F., Gemperline, D., **Caragea, C.**, Wu, F., Dobbs, D., and Honavar, V. **Prediction of RNA-Protein Interfaces Using Structural Features.** Midwest Symposium on Computational Biology and Bioinformatics, Poster Program, Chicago, Illinois (MSCBB 2007)

Gleeson, C., Hamilton, M., Lee, J.-H., **Caragea, C.**, Honavar, V., and Dobbs, D. **Generating Models as a Platform for Comparing Functional and Structural Elements of Telomerase.** The 16th Annual Growth Factor and Signal Transduction Conferences, Ames, Iowa (GFST 2007)

**Caragea, C.**, Sinapov, J., Terribilini, M., Dobbs, D., and Honavar, V. **Assessing the Performance of Macromolecular Sequence Classifiers.** Accepted to the 15th Annual International Conference On Intelligent Systems For Molecular Biology and the 6th European Conference on Computational Biology, Poster Program, Vienna, Austria (ISMB/ECCB 2007)

Caragea, D., **Caragea, C.**, Bao, J., and Honavar, V. **Ontology-Based Information Integration Using INDUS: Intelligent Data Understanding System.** Presented to Emerging Technologies Conference, Ames, Iowa (ETC 2007)

**Caragea, C.**, Sinapov, J., Silvescu, A., Dobbs, D., and Honavar, V. **Glycosylation Site Prediction using Machine Learning Approaches.** Accepted to The Eleventh Annual International Conference on Research in Computational Molecular Biology, Poster Program, Oakland, California (RECOMB 2007)

**Caragea, C.**, Sinapov, J., Silvescu, A., Dobbs, D., and Honavar, V. **Learning Classifiers to Predict Glycosylation Sites in Proteins.** Accepted to The 4th Annual Rocky Mountain Bioinformatics Conference, Poster Program, Aspen/Snowmass, Colorado (Rocky 2006)

El-Manzalawy, Y., **Caragea, C.**, Dobbs, D., Honavar, V. **On the Quality of Motifs for Protein Phosphorylation Site Prediction.** Accepted to The 14th Annual International Conference On Intelligent Systems For Molecular Biology, Poster Program, Fortaleza, Brazil (ISMB 2006)

El-Manzalawy, Y., **Caragea, C.**, Dobbs, D., Honavar, V. **Machine Learning versus Profile-Based Methods for Protein Phosphorylation Site Prediction.** Accepted to The Sixth Annual Joint Bioinformatics Symposium, Student Abstract and Poster Program, Ames, Iowa

**Caragea, C.**, Caragea, D. and Honavar, V. **Learning Support Vector Machine Classifiers from Distributed Data Sources.** In Proceedings of the Twentieth National Conference on Artificial Intelligence, Student Abstract and Poster Program, Pittsburgh, Pennsylvania. Pp. 1602-1603. AAAI Press (AAAI 2005)

## Class Projects:

**Term project for ComS 572: Principles of Artificial Intelligence**
**Text Mining Application using Tree Augmented Bayesian Networks** (in collaboration with Flavian Vasile). Design and Implementation of a tree augmented Bayesian Network, with application to the classification of news documents.

**Term project for ComS 573x: Machine Learning**
**Optimal Bayes Classifiers using Support Vector Machines Hyperplanes.** Design and implementation of a new weighting scheme for hyperplanes generated with Support Vector Machine. Comparison with the results obtained in the paper "Towards Simple, Easy-to-Understand, but Accurate Classifiers" by Caragea, D., Cook, D. and Honavar, V.

**Term project for ComS 631: Computational Complexity**
**The Complexity of Learning Bayesian Networks.** Formal definition of Bayesian Networks. Learning Bayesian Networks is NP-Hard when the size of data is small as well as when the size of data is large.

**Term project for ComS 561: Principles of Database Systems**
**DNA Microarray Experiment Database System,** a system implemented to upload microarray data resulted from biological experiments into a database, to delete data no longer needed, to modify data according to new findings, to query the data based on the project columns, to export data to a file that can be used for analysis using R.

**Term project for ComS 544: Introduction to Bioinformatics**
**Phosphorylation Sites Prediction Using SVM Classifiers** (in collaboration with Yasser El-Manzalawy).

## Teaching Experience:

**Teaching Assistant, Department of Computer Science, Iowa State University**

ComS 104: **Introduction to Computers** *Fall 2003*
ComS 104: **Introduction to Computers** *Spring 2004*
ComS 472/572: **Principles of Artificial Intelligence** *Fall 2004*
ComS 228: **Introduction to Data Structures** *Spring 2005*
ComS 330: **Discrete Computational Structures** *Fall 2005*
ComS 311: **Design and Analysis of Algorithms** *Spring 2006*

**AI Seminar, Department of Computer Science, Iowa State University**

Gave presentations, helped organizing the AI Seminar coordinated by Prof. Vasant Honavar *Spring2004-Fall2006*

**Teaching Assistant, Bioinformatics and Computational Biology Summer Institute, Iowa State University**

Module F: **Machine Learning Methods in Bioinformatics** *Summer 2006*
Module G: **Computational Analysis of "Omics" data: Machine Learning Approaches** *Summer 2007*

## Research Assistantships:

**Graduate Research Assistant**, Computer Science Department, Iowa State University. I worked in Machine Learning under Dr. Vasant Honavar. *August 2007 - May 2008, Summer 2007, August 2006 - May 2007, Summer 2006, Summer 2005*

**Graduate Research Assistant**, Ames Lab, Iowa State University. I worked in Parallel Computing under Dr. Masha Sosonkina. *Summer 2004*

## Work Experience:

I worked at **Balint Export Import**, a customs company where I administrated the database containing information about incoming cargo. I also maintained the program that computed custorms taxes for the cargo. *September 1998-June 2003*

I worked at **Quartz Computer** as database administrator. *August 1997-August 1998*