Graduate Studies in Computer Science at Dalhousie University

Evangelos Milios
Faculty of Computer Science
Dalhousie University

www.cs.dal.ca/~eem
Dalhousie U. Facts

• Founded in 1818
• The smallest Medical/Doctoral university in Canada
  – Medical school
  – Law and Business schools
  – Engineering
• World class
  – Oceanography
  – Biology
  – Medicine
  – Sciences
• Member of the G-13 research intensive universities in Canada
• Regional Research Hub for Atlantic Canada
Faculty of Computer Science
Faculty of Computer Science

• Established in 1997

• Strengths in:
  – Information retrieval, text mining
  – Health informatics & Knowledge management
  – Bioinformatics
  – Human-computer interaction, visual computing
  – Computer networks, network management, intrusion detection
  – Algorithms, graph theory, parallel computation
Interdisciplinary outlook

• Master’s degrees in:
  – Computer Science
  – Health informatics (with Medicine)
  – Electronic commerce (with Business and Law)
  – Bioinformatics (with Biology)

• Joint research projects with
  – Mathematics
  – Engineering
  – Medicine
  – Business
  – Biology
Coursework

• Number of courses depends on the degree program

• Breadth requirement must be satisfied by both Master’s and PhD students
  – For PhD students, all courses taken for a Master’s degree count
Breadth bubble diagram

**Intelligent Information Systems**
- 6403 ADV. INFO. RETRIEVAL
- 6405 DATA MINING AND WAREHOUSING
- 6509 ADV. NAT. LANG. PROCESSING
- 6707 KNOWLEDGE MANAGEMENT

**AI Fundamentals**
- 6501* INTELLIGENT SYSTEMS
- 6505 MACHINE LEARNING
- 6506 GENETIC ALGORITHMS AND PROG.
- 6508 NEUROCOMPUTING
- 6514* SEARCH AND OPTIMIZATION

**Theory**
- 6101 ADV. ANALYSIS OF ALG.
- 6102* COMPUTATIONAL GEOMETRY
- 6103* NETWORK RELIABILITY
- 6104* ALG. AND DATA STRUCT. FOR MASSIVE DATA SETS

**Bioinformatics**
- 6801 COMPUTATIONAL BIOLOGY
- 6802 BIOINFORMATIC ALGORITHMS

**Human Centric Computing**
- 6304* VISUAL PROGRAMMING
- 6306* PROGRAM COMPREHENSION
- 6406 VISUALIZATION
- 6602* DIGITAL IMAGE PROCESSING
- 6604* ADV. COMPUTER GRAPHICS
- 6606 HUMAN FACTORS IN INFO. SYS.
- 6608 ADV. COMPUTER ANIMATION

**Distributed and Software Systems**
- 6301* SW REQUIREMENT ANALYSIS/ SPEC.
- 6302* SW DEVELOPMENT AND DESIGN
- 6404* WEB INFORMATION SYSTEMS
- 6702* PARALLEL COMPUTING
- 6401 DISTRIBUTED DATABASES
- 6704 ADV. COMPUTER NETWORKS
- 6706* NETWORK DESIGN AND MANAGEMENT

**Note:**
* Denotes a course NOT offered in the current academic year.
Breadth Requirement

- ONE course from FOUR different research areas of the breadth bubble diagram
- Only courses with a CSCI number may contribute
- OUTSIDE of the above FOUR courses
  - Up to TWO grad courses from another discipline, with prior approval
  - \# of 4\textsuperscript{th} year CSCI courses +
    \# of grad courses from another discipline ≤ 2
Research overview
Research snippets
INTELLIGENT INFORMATION SYSTEMS
Web Page Categorization Using PCA

Michael Shepherd, Carolyn Watters, Jack Duffy

Web Information Filtering Lab (www.cs.dal.ca/wifl)

Recall and Precision > 0.80
Data Mining on Outlier Detection (OD) for High-Dimensional Data Streams
Q. Gao, H. Wang

• Develop innovative OD solutions based on projected outlier subspace analysis
• OD for high-dimensional data
• OD for stream data
• Research group link: http://flame.cs.dal.ca/~opami/
Visual Semantic Computation
Q. Gao, D. Gorodnichy

• Develop perceptual query language and interface toolkit for visual semantic computing
• Content based image/video retrieval
• Motion analysis for surveillance
• Generic image segmentation for supporting semantic interpretation
• Research group link: http://flame.cs.dal.ca/~ipami/
Authorship Attribution using Character N-grams
Vlado Keselj
NICHE Research Group
(kICnowledge Intensive Computing for Healthcare Enterprises)
Raza Abidi
Research Focus is Interdisciplinary

- **Computer Science**
  - Knowledge management
    - Semantic Web & Ontologies
  - Intelligent personalization
    - Semantic web service composition
    - Dynamic context-sensitive information (content) personalization

- **Health Informatics**
  - Clinical decision support systems
  - Health knowledge modeling
    - Clinical practice guidelines
    - Clinical pathways
  - Knowledge translation
  - Health data mining
Key Health Informatics Projects

• Knowledge translation in pediatric pain
  – Web 2.0, Social network analysis
• Point-of-care decision-support system for breast-cancer follow-up
  – Semantic web, Reasoning engines
• Care planning for prostate cancer through Care Maps
  – Semantic web, planning systems
• Glaucoma detection from optic discs analysis
  – Data mining, Image analysis
• Knowledge sharing patterns in Emergency Department
  – Knowledge management
• Personalized patient educational program for cardiovascular diseases
  – Adaptive hypermedia, AI

Health Informatics Research Landscape
Knowledge Morphing

“The intelligent and autonomous fusion/integration of contextually, conceptually and functionally related knowledge objects that may exist in different representation modalities and formalisms, in order to establish a comprehensive, multi-faceted and networked view of all knowledge pertaining to a domain-specific problem.”
AdWISE: Adaptive Web Information and Services Environment

- Intelligent Content Personalization
  - AI Techniques
  - IR Techniques

- Applications
  - Personalized music playlists
  - Personalized news items
  - Personalized cardiovascular risk management recommendations
Adaptive Personalized Care Planning via a Semantic Web Framework

- CarePlan is a rich temporal, process-centric, patient-specific clinical pathway that manages the evolving dynamics of a patient to meet the patient’s needs, institutional workflows and medical knowledge.
Decision Support Systems

- Semantic Web Approach
  - Knowledge Modeling
    - Ontologies
  - Knowledge Execution
    - Ontology based (logical) decision rules
    - Logic based proof engines
    - Trusted Solutions
Desktop of the future

E. Milios
Automatic Topic Extraction

E. Milios

<table>
<thead>
<tr>
<th>topic 1</th>
<th>topic 2</th>
<th>topic 3</th>
<th>topic 4</th>
<th>topic 5</th>
<th>topic 6</th>
<th>topic 7</th>
<th>topic 8</th>
<th>topic 9</th>
<th>topic 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>error</td>
<td>neuron</td>
<td>image</td>
<td>analog</td>
<td>data</td>
<td>control</td>
<td>function</td>
<td>rule</td>
<td>distribution</td>
<td></td>
</tr>
<tr>
<td>generalization</td>
<td>neurons</td>
<td>images</td>
<td>circuit</td>
<td>clustering</td>
<td>model</td>
<td>functions</td>
<td>rules</td>
<td>probability</td>
<td></td>
</tr>
<tr>
<td>learning</td>
<td>synaptic</td>
<td>object</td>
<td>current</td>
<td>principal</td>
<td>motor</td>
<td>basis</td>
<td>set</td>
<td>gaussian</td>
<td></td>
</tr>
<tr>
<td>training</td>
<td>firing</td>
<td>recognition</td>
<td>figure</td>
<td>cluster</td>
<td>forward</td>
<td>linear</td>
<td>step</td>
<td>data</td>
<td></td>
</tr>
<tr>
<td>optimal</td>
<td>spike</td>
<td>face</td>
<td>chip</td>
<td>pea</td>
<td>inverse</td>
<td>regression</td>
<td>form</td>
<td>parameters</td>
<td></td>
</tr>
<tr>
<td>order</td>
<td>time</td>
<td>objects</td>
<td>voltage</td>
<td>set</td>
<td>dynamics</td>
<td>kernel</td>
<td>fuzzy</td>
<td>model</td>
<td></td>
</tr>
<tr>
<td>large</td>
<td>activity</td>
<td>hand</td>
<td>vlsi</td>
<td>algorithm</td>
<td>controller</td>
<td>space</td>
<td>problem</td>
<td>bayesian</td>
<td></td>
</tr>
<tr>
<td>average</td>
<td>rate</td>
<td>pixel</td>
<td>circuits</td>
<td>points</td>
<td>feedback</td>
<td>gaussian</td>
<td>relative</td>
<td>mixture</td>
<td></td>
</tr>
<tr>
<td>small</td>
<td>synapses</td>
<td>system</td>
<td>digital</td>
<td>approach</td>
<td>system</td>
<td>approximation</td>
<td>extraction</td>
<td>density</td>
<td></td>
</tr>
<tr>
<td>examples</td>
<td>potential</td>
<td>view</td>
<td>implementation</td>
<td>clusters</td>
<td>position</td>
<td>rbf</td>
<td>expert</td>
<td>likelihood</td>
<td></td>
</tr>
</tbody>
</table>

Figure 2. Example word-topics for the NIPS dataset

<table>
<thead>
<tr>
<th>topic 1</th>
<th>topic 2</th>
<th>topic 3</th>
<th>topic 4</th>
<th>topic 5</th>
<th>topic 6</th>
<th>topic 7</th>
<th>topic 8</th>
<th>topic 9</th>
<th>topic 10</th>
</tr>
</thead>
<tbody>
<tr>
<td>language</td>
<td>game</td>
<td>church</td>
<td>house</td>
<td>air</td>
<td>league</td>
<td>war</td>
<td>apollo</td>
<td>party</td>
<td>system</td>
</tr>
<tr>
<td>english</td>
<td>player</td>
<td>god</td>
<td>parliament</td>
<td>aircraft</td>
<td>football</td>
<td>german</td>
<td>earth</td>
<td>government</td>
<td>computer</td>
</tr>
<tr>
<td>greek</td>
<td>cards</td>
<td>christian</td>
<td>members</td>
<td>world</td>
<td>team</td>
<td>army</td>
<td>moon</td>
<td>president</td>
<td>game</td>
</tr>
<tr>
<td>languages</td>
<td>players</td>
<td>jesus</td>
<td>commons</td>
<td>force</td>
<td>world</td>
<td>soviet</td>
<td>lunar</td>
<td>political</td>
<td>games</td>
</tr>
<tr>
<td>word</td>
<td>games</td>
<td>christ</td>
<td>lords</td>
<td>military</td>
<td>club</td>
<td>battle</td>
<td>time</td>
<td>national</td>
<td>apple</td>
</tr>
<tr>
<td>russell</td>
<td>play</td>
<td>orthodox</td>
<td>bill</td>
<td>ship</td>
<td>home</td>
<td>germany</td>
<td>mission</td>
<td>minister</td>
<td>atari</td>
</tr>
<tr>
<td>century</td>
<td>card</td>
<td>baptism</td>
<td>act</td>
<td>gun</td>
<td>season</td>
<td>world</td>
<td>program</td>
<td>states</td>
<td>commodore</td>
</tr>
<tr>
<td>theory</td>
<td>hand</td>
<td>life</td>
<td>power</td>
<td>war</td>
<td>won</td>
<td>forces</td>
<td>module</td>
<td>united</td>
<td>home</td>
</tr>
<tr>
<td>words</td>
<td>round</td>
<td>catholic</td>
<td>chopin</td>
<td>ships</td>
<td>game</td>
<td>french</td>
<td>jpg</td>
<td>election</td>
<td>software</td>
</tr>
<tr>
<td>modern</td>
<td>played</td>
<td>roman</td>
<td>speaker</td>
<td>navy</td>
<td>major</td>
<td>union</td>
<td>crew</td>
<td>state</td>
<td>video</td>
</tr>
</tbody>
</table>

Figure 3. Example word-topics for the Wikipedia dataset
Peer-to-Peer Document Management

V. Keselj, E. Milios, S. Abidi

A P2P network cloud hosting various users that share their knowledge resources. Each user has a set of profiles and document corpora.

A single user's view of the available knowledge resources in a P2P environment.

User's document corpora linked to an organizational semantic model.
Experience Management
E. Milios, N. Zincir-Heywood
AI FUNDAMENTALS
Computational Neuroscience

Dr. Thomas Trappenberg

Machine Learning

Reinforcement Perceptron

SOM

Sensory  Evaluative  Motor feedback
Genetic Programming

Problem Decomposition

Co-evolutionary behaviors

Evolving Computer Programs

Game Strategy Learning

Malcolm Heywood
Evolutionary Computation

- Evolutionary algorithms are optimisation strategies “gleaned from nature”
- Areas of application range from engineering design and control to financial forecasting and art
- Research of Dalhousie’s Evolutionary Computation group focuses on understanding, improving, and developing adaptive strategies
- Contact: Dr. Dirk Arnold (http://www.cs.dal.ca/~dirk)
THEORY
Disk I/O bottleneck when processing massive datasets

Low cache efficiency in traditional algorithms

Need algorithms with high access locality to
  - Take advantage of caches
  - Take advantage of disk read-ahead

Techniques fundamentally different from traditional algorithms!
Algorithms and Data Structures for Memory Hierarchies

- **Geometric problems**
  - Databases (range queries, etc)
  - GIS (map overlay, window queries, etc)
  - ...

- **Graph problems**
  - Web modeling
  - GIS (route planning, logistics)
  - Bioinformatics (protein clustering, etc)
  - ...

Norbert Zeh
Canada Research Chair in Algorithms for Memory Hierarchies
Fault-tolerant networks

• Design and Reconfiguration of fault-tolerant networks.
• Objectives: construct a network that
  – Can continue to operate in the presence of certain faults,
  – Is optimal or near-optimal in cost,
• Cost will depend on the parameters to be optimized
• Efficiency of reconfiguration measured by the time needed to identify a healthy sub-graph of the network (that excludes the defective components).
BIOINFORMATICS
Bio-informatics

Optimizing confidence intervals in phylogeny
Parallel Computing in protein phylogeny
Sequence alignment curation using Artificial Intelligence
A C++ bioinformatics library
Interactive Phylogeny
Protein Biophysics and the substitution process
**Structural Evolution**
Folding of protein loops

Dr. Christian Blouin
HUMAN CENTRIC COMPUTING
Visualisation in software development

- Visual Languages (VL)
  - graphical notations that directly express the multidimensional structure of algorithms and data.
- Visualisation of execution
- End-user and domain-specific programming

Some current projects

- Design of structured objects
- Programming by demonstration
- VLs for industrial software development
- Spreadsheet programming and templating
  - Example: Gaussian elimination for solving sets of linear equations (not a typical usual end-user application!)...
Applying a template

- select the template to apply - gauss
- select arrays in the worksheet corresponding to the parameters of gauss
- outlines turn green when shapes are acceptable
- click the “apply” button
The Dalhousie Graphics and Visualization Lab
The Graphics and Visualization Lab

• The focus is on both:
  – the development of new graphical techniques, and
  – the application of those techniques, often in cross-disciplinary areas

• Our lab incorporates expertise in areas such as:
  – image processing
  – 3D computer graphics
  – physically-based rendering
  – visualization
  – and, traditional art
Graduate Courses & Faculty Members

• Visualization (6406)
  – focuses on graphical techniques for data visualization that assist in the extraction of meaning from datasets

• Advanced Computer Animation (6608)
  – covers topics in computer animation, including forward and inverse kinematics, motion capture, and physically based modelling

• Digital Image Processing (6602)
  – covers topics in digital picture processing such as visual perception, digitization, compression and enhancement
DISTRIBUTED AND SOFTWARE SYSTEMS
Network Information Management and Security

Nur Zincir-Heywood

www.cs.dal.ca/~zincir

Network Analysis

Attack Modelling

Fault Management

Computer Security
Collaborative User Services for Private Data Management (CUSP)

The CUSP (Collaborative User Services for Private Data Management) project intends to deliver sophisticated user privacy services over the Semantic Web. This Canadian project is a collaborative effort between faculty in the Sobey School of Business, Saint Mary’s University and the Faculty of Computer Science, Dalhousie University.

Currently many knowledge-intensive privacy-related tasks are manual. Using Semantic Web technologies (OWL, RDF, XML, UDDI, SOAP, and WSDL), knowledge-base and database methodologies, and building on the P3P platform (XML vocabulary for privacy), the CUSP project automates human decision making processes with respect to online privacy.

Further information at http://users.cs.dal.ca/~bodorik/Cusp.htm

Peter Bodorik
www.cs.dal.ca/~bodorik
This project provides technologies to support compliance to privacy regulations in a Web Services Architecture. Automated agents examine messages exchanged when invoking web-services. The agents utilize a Privacy Knowledge Base to ensure that Private Information that is exchanged satisfies applicable privacy policies.

For further information contact Dr. Bodorik or Dr. Jutla.
Highly Scalable High Performance Caching Architecture
Achieved by Interoperable Cache Managers and Data Servers

DB servers are becoming bottlenecks in enterprise caching architectures. A highly scalable and high performance caching architecture is achieved by

- Offloading the caching responsibilities of a DB server to Global Cache Managers (GCMs)
- Local Cache Managers (LCMs) coordination with Cache Data Servers in caching protocols
- Interoperable caching protocols that support applications with different caching requirements

For further information contact Dr. Bodorik at www.cs.dal.ca/~bodorik

Peter Bodorik
www.cs.dal.ca/~bodorik
CSCI 6401 Distributed Databases

Instructor:  Peter Bodorik

www.cs.dal.ca/~bodorik;  email:  bodorik@cs.dal.ca

Mondays,  Wednesdays  11:05-12:25,  Computer Science LAB-3

Objectives

The main objective of this course is to examine the issues arising in the design and implementation of distributed databases. Another objective is to examine current developments in the use of DBs and information systems in support of Enterprise Information Systems.

Course Organization

A portion of the course is devoted to the subject matter appearing in the textbook. Lectures are used to outline the problems and their solutions. You are expected to study the subject matter and pass assignments and tests.

You will investigate an assigned topic dealing with usage of DBs or systems accessing DBs, give a presentation on it and submit a report.
WISE (Wireless Security) Group

- Investigate protocol vulnerabilities in wireless networks – WiFi, WiMAX and Ad Hoc Wireless
- Build a manual for best practice for wireless security.
- Design intrusion detection and prevention mechanisms for enhancing security.
- Implement prototypes and build a test bed for validating these detection and prevention mechanisms.
- Integrate security and quality of service in heterogeneous and hybrid networks.
WISE (Wireless Security) Group

Applications

- Voice
- Video
- Security
- Wireless Network
- Guaranteed Quality of Service
- Data
- Image
Graduate School Information
Choosing advisor & thesis topic

From: *How to succeed in graduate school* (by Marie deJardins, SRI International)

- a good thesis topic is interesting:
  - to you,
  - to your advisor, and
  - to the research community

- Professors may have
  - Well defined long-term research programs and expect their students to contribute directly
  - Much looser, but still related ongoing projects.
  - Tendency to take on anyone with an interesting idea (beware of advisor lack of commitment)
Scope of reading & topic

• Awareness & Reading
  – Be selective: you'll never be able to read everything that might be relevant
  – Become and stay aware of directly related research

• Topic options
  – Narrow, well defined topic.
    • Plus: finish fast
    • Minus: it may not be as exciting
  – Exotic topic
    • Plus: potentially exciting
    • Minus: difficulty convincing people it's worthwhile.
Good topic choices

• Solve a real problem, not a toy problem
• Choose:
  – a central problem that's solvable and acceptable
  – with extensions and additions that:
    • are successively riskier and that
    • will make the thesis more exciting.
Programme Form

• FGS is responsible for the program of all graduate students at Dalhousie.
• Coursework for a graduate student is approved by a faculty advisor
• Programme Form
  – Shows the list of approved courses for a student
  – A contract between the student and Dalhousie
  – List can be changed later (with approval)
**Graduate Student Programme Form**

**To be completed by department one month after registration.**

**Student Name:**

**Address:**

**Telephone:**  
**E-mail:**

**Student Number:**  
**Start Date:**

**Years of Programme Fee Requirement:**  
- [ ] 1 Year  
- [ ] 2 Years  
- [ ] 3 Years  
- [ ] Fee per Course Programme

**Degree Programme:**

**Registration Status:**  
- [ ] Full-Time  
- [ ] Part-Time  
- PhD Students must register Full-Time

**Department/School:**

**Area of Specialisation:**

**Total of full-credit classes (with credit value) required for degree (excluding thesis):**

**Indicate all requirements; include both course number and credit value**

**Required Courses for Programme**

<table>
<thead>
<tr>
<th>Subject Code</th>
<th>Course Number</th>
<th># Credit Hours</th>
<th>Subject Code</th>
<th>Course Number</th>
<th># Credit Hours</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Credits to be added later:**

**Credits Audited:**

**Ancillary Credits (e.g., undergraduate courses not required to complete the programme):**

**Additional Requirements for Degree (give class number if applicable):**

- [ ] Thesis:  
- [ ] Project:  
- [ ] Practicum:  
- [ ] Language Requirement(s):

**Other Requirements (please specify):**

**Name(s) of supervisor(s):**

**Names of committee members (if known):**

**Approvals:**

**Student:**  
**Date:**

**Supervisor:**  
**Date:**

**Graduate Coordinator:**  
**Date:**

**Faculty of Graduate Studies:**  
**Date:**

*Revised September 2003*
For more information

- WWW: http://www.cs.dal.ca/graduate/
- Email: grad@cs.dal.ca
- Resources about graduate school:
  - thesis writing
  - how to do research
  - how to give presentations
  - job interview preparation
    http://users.cs.dal.ca/~eem/gradResources/gradResources.htm