Graduate Studies in Computer Science at Dalhousie University

Evangelos Milios
Faculty of Computer Science
Dalhousie University
www.cs.dal.ca/~eem
Bird’s eye view of Halifax
Halifax Fun
Halifax, Nova Scotia

- Northernmost harbour that does not freeze in the winter
- Relatively mild climate
- Metropolis of Atlantic Canada (incl. Nova Scotia, New Brunswick, Prince Edward Island, and Newfoundland)
- Regional economic, cultural and research hub
- Settled by Europeans in the 18th century
Dalhousie U. Facts

• Founded in 1818
• The smallest Medical/Doctoral university in Canada
  – Medical school
  – Law school
  – Engineering
  – Business school
• World class
  – Oceanography
  – Biology
  – Medicine
  – Sciences
• 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
  – 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
Research snippets
Networked Information Spaces:

Modelling and Mining
Documents are networked into information spaces

- World Wide Web
- Blog space
- Scientific and Medical Literature
- Patents
- Common Law
Desktop of the future

- Personal Document Library
  - Metadata extraction
  - Keyterm extraction
  - Key sentence extraction - Summarization
  - Lucene Apache Search Engine Inverted Index
    - Taxonomies: Hierarchical clustering
    - Query
      - Browsing
    - Organization: clustering, ranking, informative snippets, keywords, key sentences
    - Results
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.

User's document corpora linked to an organizational semantic model.

A single user's view of the available knowledge resources in a P2P environment.
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>
</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>
</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>
</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>
</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>
</tr>
<tr>
<td>optimal</td>
<td>spike</td>
<td>face</td>
<td>chip</td>
<td>pca</td>
<td>inverse</td>
<td>regression</td>
<td>form</td>
<td>parameters</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>
</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>
</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>
</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>
</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>
</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
Experience Management
E. Milios, N. Zincir-Heywood
Connectivity of the Citation graph

J. Janssen, E. Milios

(a) 68.5% of the nodes have no incoming link

(b) 58% of the nodes in the giant Weakly Connected Component (WCC) account for a big Biconnected Component (BCC)

papers inside this area haven’t been cited yet.

papers inside this area have been cited.

papers form a biconnected nucleus, it takes 58%.
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
Authorship Attribution using Character N-grams

Vlado Keselj
Dickens: Christmas Carol

Dickens: A Tale of Two Cities

Carroll: Alice’s adventures in wonderland
Network Traffic Classification
Nur Zincir-Heywood

1st Level
- Duration
- Protocol
- Service
- Status Flag
- Destination bytes
- Source bytes

2nd Level
- #Attack Connections
- NIMS
  - http://www.cs.dal.ca/projectx

3rd Level
- #Normal Connections
Enhancing Face-to-Face Collaboration
EDGE Lab
Dalhousie University
Dr. Kori Inkpen

Exploring effective interaction techniques and input devices for rich face-to-face environments
Aware Home
Jacob Slonim
Visual Computing & Design

Phil Cox

• The role of visualisation in software development
  – Visual programming languages (VPL)
  – Visualisation of execution
  – End-user and domain-specific programming

• Some projects
  – Design of structured objects
  – Programming by demonstration
  – VPLs for industrial software development
  – Spreadsheet programming and templating

• Example: Gaussian elimination for solving sets of linear equation (not a typical usual end-user application)
### Applying a template

- contents of solution vector (formulae) are computed, and evaluated
- select array corresponding to parameter `gauss`
- outlines are acceptable
- click the "apply" button
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
Computational Neuroscience

Machine Learning & Data Mining

eCommerce

Dr. Thomas Trappenberg
NICHE Research Group (kNowledge 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
The Dalhousie Graphics and Visualization Lab

GVLAB
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 Graphics (6604)
  – covers topics in computer graphics, including rendering, geometric modeling, and computer animation

• Digital Image Processing (6602)
  – covers topics in digital picture processing such as visual perception, digitization, compression and enhancement
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)
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).

Zizo Farrag
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!
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
Graduate School Information
Admission requirements

- Grade point average at least 3.7 (on a 4.3 scale)
- Strong reference (recommendation) letters
- Publications highly desired (for the PhD program)
- TOEFL
- GRE (optional)
Application process

• Visit: [http://www.cs.dal.ca/graduate/](http://www.cs.dal.ca/graduate/)

• **Deadline**: January 10 (for September)

• Students need support from a prospective **supervisor** to be admitted
  
  – Feel free to contact faculty members in your areas of interest two weeks after you have sent your application
  
  – Acceptable to switch supervisors after admission
  
  – Minimal delay if done within the 8 months
How much money do I need?

<table>
<thead>
<tr>
<th>EXPENSES</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuition+health ins.</td>
<td>8206</td>
</tr>
<tr>
<td>Housing+food</td>
<td>8400</td>
</tr>
<tr>
<td>Personal+books</td>
<td>2100</td>
</tr>
<tr>
<td><strong>-- total</strong></td>
<td><strong>18706</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SUPPORT</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Scholarship</td>
<td>6151</td>
</tr>
<tr>
<td>Research assistantship</td>
<td>9324</td>
</tr>
<tr>
<td>Teaching assistantship</td>
<td>3248</td>
</tr>
<tr>
<td><strong>-- total</strong></td>
<td><strong>18723</strong></td>
</tr>
</tbody>
</table>
How much money do I need?

- Cost of living differs among Canadian cities.
- $20,000 in Halifax is the same as:

<table>
<thead>
<tr>
<th>City</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>Edmonton</td>
<td>23800</td>
</tr>
<tr>
<td>Vancouver</td>
<td>31000</td>
</tr>
<tr>
<td>Toronto</td>
<td>32600</td>
</tr>
<tr>
<td>Hamilton</td>
<td>22300</td>
</tr>
<tr>
<td>Montreal</td>
<td>30000</td>
</tr>
<tr>
<td>Ottawa</td>
<td>30600</td>
</tr>
</tbody>
</table>

Data from: [http://www.usask.ca/cgsr/comparison.php](http://www.usask.ca/cgsr/comparison.php)
How to choose a 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)
How to choose a thesis 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.
How to choose a thesis topic?

• 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.
For more information

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