NSERC CRSNG *Investing in people, discovery and innovation*
Investir dans les gens, la découverte et l'innovation

FORM 200
Application for an NSERC Scholarship or Fellowship
COVER PAGE

Date: 2002/09/26

Family name of applicant	Given name	Initial(s) of all given names	Personal identification no. (PIN)
Foss	Andrew	APO	

ADDRESSES

Current address	Permanent mailing address (if different than current address)
10610 79 Ave, Appt. 202 Edmonton, AB CANADA T6E 1S1	
If current address is temporary, indicate leaving date	Telephone number at permanent mailing address

Telephone number	Facsimile number	E-mail address
1 (780) 486-0765	(780) 486-0765	afoss@cs.ualberta.ca

CITIZENSHIP

[ ] Canadian citizen [X] Permanent resident of Canada [ ] Other

Indicate date of landing as per Form IMM 1000	Indicate country of citizenship
2001/11/15	UK

LANGUAGE OF CORRESPONDENCE

I wish to receive my correspondence in:

[X] English [ ] French

SIGNATURE

I hereby agree that any award made to me as a result of this application will be subject to the general conditions governing scholarships and fellowships. These conditions are outlined in this Web site in the NSERC *Program Guide for Students and Fellows,* the *Visiting Fellowships in Canadian Government Laboratories* guide, and the description of the NATO Science Fellowships program.

Applicant's signature

Personal information collected on this form and appendices will be stored in the Personal Information Bank for the appropriate program.

NSERC CRSNG *Investing in people, discovery and innovation*
Investir dans les gens, la découverte et l'innovation

FORM 200
Application for an
NSERC Scholarship or Fellowship

Date: 2002/09/26

Family name of applicant	Given name	Initial(s) of all given names	Personal identification no. (PIN)
Foss	Andrew	APO	

ACADEMIC BACKGROUND (include only current and past degree programs)

Degree	Name of discipline	Department, institution and country	Month and year started	Month and year awarded/expected
Bachelor's	Physics	Physics Oxford University, UK	9/1972	6/1976
Master's	Computing Science	Computing Science Alberta, CANADA	9/2000	11/2002
Doctorate				

ACADEMIC, RESEARCH AND OTHER RELEVANT WORK EXPERIENCE

Position held and nature of work (begin with current)	Organization and department	Supervisor	Period (mm/yyyy-mm/yyyy)
Research Assistant Research on clustering (part-time)	University of Alberta Computing Science	Dr Osmar Zaiane	05/2001 -12/2001
Self-employed Consulting, software design, development, marketing and support (full-time)	ShriSource		01/1995 -11/2001
Teaching Assistant Lab supervision, student orientation, guidance and marking assignments (part-time)	University of Alberta Computing Science	Dr Eleni Stroulia	01/2001 -04/2001
Various Developing research labs, reviewing and managing publication of research, lecturing (full-time)	Maharishi European Research University Physics and Social Sciences	Dr G Clements, Dr B Rigby and others	01/1976 -12/1994
Research Assistant Observations, programming, data capture and processing, other research activities (full-time)	Royal Greenwich Observatory	Dr Paul Lee	02/1972 -09/1972

Personal identification no. (PIN)	Family name, given name and initial(s) of applicant
	Foss, Andrew APO

AWARD APPLIED FOR

Type of award	Proposed starting date of award
Postgraduate Scholarships - PGS B	2003/01/01

Proposed degree program (e.g. Masters, Doctorate)	Proposed field of study/research	Research subject code
Doctorate	Unsupervised clustering (Computing Science - Databases)	2705

Title of proposed research (not required for VF applicants)

Unsupervised Clustering of Very Large Databases

List ten (10) key words that describe your proposed research. Use commas to separate them. (Not required for VF applicants.)

data clustering, data mining, outlier detection, high dimensionality, scalable clustering algorithm, very large databases, forecasting, pattern recognition, unsupervised classification, noise filtering

PROPOSED LOCATION(S) OF TENURE (in order of preference)

Institution/organization	Department	Proposed supervisor
Alberta	Computing Science	Dr Osmar Zaiane

SECTION TO BE COMPLETED BY PGS APPLICANTS ONLY

Indicate the number of months of graduate studies (master's and doctoral) you will have completed as of December 31 of the year of application.

24 months of full-time graduate studies

0 months of part-time graduate studies

Indicate if you are attending university at the time of application.

[ ] Attending part time [X] Attending full time [ ] Not attending

Are you applying for tenure of your award abroad?

[ ] Yes [X] No

Personal identification no.(PIN)	Family name, given name and initial(s) of applicant
	Foss, Andrew APO

THESES COMPLETED OR IN PROGRESS

1. Degree	Supervisor	Date degree requirements completed
MSc	Dr Osmar Zaiane	09/2002

Title of thesis
Automatic Categorical Data Clustering and Spatial Data Clustering by Consecutive Resolution Refinement

2. Degree	Supervisor	Date degree requirements completed

Title of thesis

SUMMARY OF THESIS MOST RECENTLY COMPLETED OR IN PROGRESS (honours project, master's or doctoral)

Use plain language. Do not reproduce abstract of thesis.

The first step in finding useful information from a very large database (a.k.a data mining) is to identify the patterns in the data. The first stage of this is called clustering. A great deal of work has been done on this and yet no solution fully addresses the need for an automatic process that scales to handle very large numbers of data points and high dimensionality. Virtually every clustering algorithm demands parameters from the user and if these are not well tuned to the data set, the output is meaningless. However, in most cases, the user will not know in advance what parameters to set. This thesis set out to investigate automatic clustering - unsupervised classification - of large data sets both in the categorical and spatial data arena.

New algorithms (TURN, TURN*) were developed and formalised that succeed very well in this, TURN clustering categorical data better than a leading alternative and TURN* clustering large spatial data sets faster and, in almost every case, more precisely than any of the six other leading contenders from different categories of approach to this problem. TURN* was also the only algorithm that could find the correct answer without parameter input and showed near linear scaling with data set size while handling arbitrary cluster shapes. Furthermore it was successful in handling very noisy and complex data sets on which, for example, the most popular clustering approach - k-means - and its derivatives fail. It was shown formally that the method is insensitive to input data order and the clusters found are maximal. The approach is closest to density based approaches and its automation algorithm represents a significant advance in cluster validation, a field which has so far largely failed to develop beyond handling spherical shaped clusters.

An important feature of the TURN* algorithm is it's ability to detect automatically the most 'interesting' resolution levels - the ones that a human would tend to identify due to their inherent stability. This is the first time this has been achieved.

The work done so far also offers the basis for a complete clustering solution that could offer the potential for finding quite complete information including clusters embedded within clusters, clusters subject to constraints like walls and bridges in a GIS dataset, and outliers in a more robust way than current approaches.

Personal identification no.(PIN)	Family name, given name and initial(s) of applicant
	Foss Andrew APO

Justification for location of tenure

Provide a rationale for your choice(s) for location of tenure (maximum 1 page). See instructions for further details.

I have chosen to continue with Dr Osmar Zaiane and the University of Alberta database lab for the following reasons. The Computer Science department at the University of Alberta is one of the best funded and most respected in the country. The department is hiring aggressively and recently recruited Dr. Joerg Sander who has developed the clustering algorithm most closely related to mine. His presence is invaluable even though he is not my supervisor.

While many professors have only a few graduate students, Dr Zaiane has 10 or more and there is always a waiting list. This is because he is both a first-class supervisor and human being but also exceptionally supportive and proactive to get his students' work into print. At the next major conference (ICDM 2002, Japan), three of his students have papers accepted, including myself, and one more is awaiting the results of the review process. Just the group working with him alone constitutes a remarkable concentration of talent and ideas and I feel greatly benefited from working with and alongside them. In my first two terms of the MSc, I produced three publishable results from projects with different professors but it was Dr Zaiane that actively encouraged me so the first paper was accepted even before I finished my course work.

NSERC CRSNG *Investing in people, discovery and innovation*
Investir dans les gens, la découverte et l'innovation

URGENT

Please return to:

By (date):

APPENDIX 1
Report on the Applicant
NSERC Scholarship or Fellowship
(FORM 200)

Read the instructions before you complete this report. For PGS applicants, use this page only. For all other applicants, use one additional page if necessary.

In accordance with the *Privacy Act,* this report will be accessible to the applicant. This report, including your name, may also be disclosed to organizations whose fellowships are administered by NSERC and to partner agencies that offer supplements to NSERC awards.

Date: 2002/09/26

Family name of applicant	Given name	Initial(s) of all given names	Personal identification no. (PIN)
Foss	Andrew	APO	

Type of award applied for: PGS B

Size of comparison group:

Using a percentage scale, rate the applicant on each of the following three evaluation criteria (e.g., Top 5%, Top 10%, etc. or N/A – Inadequate opportunity to observe).

1. Research ability or potential
2. Oral and written communication skills
3. Interpersonal and leadership abilities

Comment on the applicant with respect to each criterion.

I have known the applicant in my capacity as _____ for _____ years.

Printed name, title, and affiliation of respondent

I have received and read a copy of the applicant's completed Parts I and II of Form 200.

_____
Signature of respondent

NSERC CRSNG *Investing in people, discovery and innovation*
Investir dans les gens, la découverte et l'innovation

URGENT

Please return to:

By (date):

APPENDIX 1
Report on the Applicant
NSERC Scholarship or Fellowship
(FORM 200)

Read the instructions before you complete this report. For PGS applicants, use this page only. For all other applicants, use one additional page if necessary.

In accordance with the *Privacy Act,* this report will be accessible to the applicant. This report, including your name, may also be disclosed to organizations whose fellowships are administered by NSERC and to partner agencies that offer supplements to NSERC awards.

Date: 2002/09/26

Family name of applicant	Given name	Initial(s) of all given names	Personal identification no. (PIN)
Foss	Andrew	APO	

Type of award applied for: PGS B

Size of comparison group:

Using a percentage scale, rate the applicant on each of the following three evaluation criteria (e.g., Top 5%, Top 10%, etc. or N/A – Inadequate opportunity to observe).

1. Research ability or potential
2. Oral and written communication skills
3. Interpersonal and leadership abilities

Comment on the applicant with respect to each criterion.

I have known the applicant in my capacity as _____ for _____ years.

Printed name, title, and affiliation of respondent

I have received and read a copy of the applicant's completed Parts I and II of Form 200.

_____
Signature of respondent

NSERC CRSNG *Investing in people, discovery and innovation*
Investir dans les gens, la découverte et l'innovation

APPENDIX 2
Departmental/University Evaluation
NSERC Scholarship or Fellowship
(FORM 200)

Read the instructions before you complete this report. In accordance with the *Privacy Act,* this report will be accessible to the applicant. This report, including your name, may also be disclosed to organizations whose fellowships are administered by NSERC and to partner agencies that offer supplements to NSERC awards.

Date: 2002/09/26

PART I: TO BE COMPLETED BY THE APPLICANT

Family name	Given name	Initial(s) of all given names	Personal identification no. (PIN)
Foss	Andrew	APO	

University (indicate if current or former): Alberta (current)

University department: Computing Science

[X] Application submitted through university

[ ] Application submitted directly to NSERC

PART II: TO BE COMPLETED BY THE HEAD OF THE DEPARTMENT OR REPRESENTATIVE

Description of peer group

Departmental ranking of applicant in comparison to peer group: _____ of _____

Comments on the applicant

[For PDF only] Ranked list of all PDF applicants from the department in the current competition

_____ Signature of head of department (or representative)

_____ Date

PART III [FOR PGS ONLY]: TO BE COMPLETED BY THE SCHOLARSHIP LIAISON OFFICER

Weighted annual averages

Second last year _____ / _____ (maximum)

Last year _____ / _____ (maximum)

University ranking of applicant

_____ of _____

Additional comments on the applicant

NSERC CRSNG *Investing in people, discovery and innovation*
Investir dans les gens, la découverte et l'innovation

APPENDIX 3
Consent Form and Statistical Information
NSERC Scholarship or Fellowship
(FORM 200)

Date
2002/09/26

Family name of applicant	Given name	Initial(s) of all given names	Personal identification no. (PIN)
Foss	Andrew	APO	

CONSENT FROM THE APPLICANT

If you wish to be considered for other types of scholarships, fellowships or supplements offered by the Canadian government, companies and/or associations as described in this Web site in the *Program Guide for Students and Fellows,* you must complete and sign this form. By completing and signing this section, you agree to have your application (including Appendices 1 and 2) disclosed to the interested institutions (see the *Access to Information Act* and the *Privacy Act,* under Policies and Guidelines in the *Program Guide for Students and Fellows,* on this Web site).

If you are eligible and wish to be considered for one or more supplements, indicate which ones, as described in NSERC 's *Program Guide for Students and Fellows.*

PDF APPLICANTS

If your application is not successful but is judged to be meritorious, would you like to be considered for the

☐ Visiting Fellowships (VF) program?

☐ Industrial Research Fellowships (IRF) program?

SIGNATURE

_____
Applicant's signature

STATISTICAL INFORMATION

Information on gender, designated groups and language capability is collected for administrative and statistical purposes and for the promotion of increased participation by designated groups in science and engineering programs. It will not be released to persons outside NSERC. For more information, see, in this Web site, the *Program Guide for Students and Fellows* and the introduction of the *Visiting Fellowships in Canadian Government Laboratories* guide.

GENDER (completion optional)	DESIGNATED GROUPS (completion optional)
☒ Male ☐ Female	☐ Aboriginal ☐ Visible minority ☐ Disabled

LANGUAGE CAPABILITY (completion optional)

	Read	Write	Speak
English	☒	☒	☒
French	☐	☐	☐

Canada