Deepak Munjal (Supervisor: Evangelos Milios), Computer Science
Date: March 08, 2019
REB file #:   2019-4699
Title: Visual Analytics of Research Community Expertise in Space and Time

To: Social Sciences and Humanities Research Ethics Board Review Letter

The review letter from the Social Sciences and Humanities Research Ethics Board has been studied thoroughly and all corrections and clarifications has been added accordingly. Each of the Board's comment is followed by a precise indication of how it has been addressed, including reference to the corresponding sections and page numbers below in this cover letter. Each response has been highlighted with a different text color.

The revised version of the submission itself has also been attached with this email. Each change has been highlighted including any removals and additions to the text.

RECOMMENDATION: Revisions and/or clarifications are requested.

GENERAL COMMENTS
At times, the application was unclear, and there is a need for major copy editing to take place on documents that are intended to be distributed to participants.

Major editing has been done throughout application and the documents that are intended to be distributed to participants. Moreover, it has been proofread and corrected for all grammatical errors.

2.3 RECRUITMENT
2.3.2 There is no information about how the PI will recruit participants (who are a range of students, staff, and faculty). How will participants be identified? Please clarify the procedure (elaborate and be more specific with details) and tools required for recruitment from the point of advertisement to when they come in to do the study.

Section - 2.3.1 and 2.3.2
Page Numbers – 7 and 8

The population for our study will be Graduate students, Postdocs and Research Assistants in Computer Science at Dalhousie University who have read an academic research paper. That research paper can be their own published paper or a published paper that they have studied as part of their thesis or courses or just for interest. We will need 10 participants to have a credible statistical analysis. This

number of participants is selected based on similar user studies performed in this area.
Recruitment Plans: PI will send an e-mail to all Graduate students, Postdocs and Research Assistants in Computer Science at csgrads@cs.dal.ca . Recruitment paper posters will also be posted on notice boards in the Goldberg Computer Science building. Interested candidates can read the email or posters, read consent form by opening the link available in the both recruitment email and recruitment poster. If they are interested, they can email to PI on the email mentioned in both email and poster. They need to send their Degree name and year of study in the email. If we receive email from more than 10 candidates who are interested in taking part in the study, then we will select only 10 participants out of those candidates based on the seniority level (Degree name and Year of study). Further, PI will reply to them and invite them for the study at a date and time based on their availability.

Recruitment Instruments/Tools: PI will send an email at csgrads@cs.dal.ca and paper posters will be used as well. These paper posters will be put on the notice boards in Goldberg Computer Science building, Dalhousie University. The email and poster is attached in appendices.

It is also not clear the purpose of the screening tool (Appendix C) if "there is no criterion based on the screening questionnaire". How will the PI identify (after recruiting) if participants are familiar with the concepts of classification? How will you know if participants are familiar with ACMDSP? Read scientific papers? How are scientific papers defined (and does the researcher mean 'academic' papers rather than 'scientific' papers)? Will experience using databases and excel be important?

Section - 2.3.1
Page Number - 7

Screening questionnaire (Previously Appendix C) has been removed as all Graduate students, Postdocs and Research Assistants in Computer Science read academic research papers as part of their courses or Thesis or Research group meetings. Familiarity with classification and ACMDSP is not required and hence removed. A research paper is defined as an academic research paper. That academic research paper can be participant's own published paper or a published paper that they have studied/read as part of their thesis or courses or just for interest. All Computer Science Graduate students, Postdocs and Research Assistants know databases and excel.

In section 2.3.1 it indicates that the study population are individuals who are working on a thesis/report or planning to write a survey paper, however, there appears to be no screening for this inclusion criteria. What does a "collection of research documents of their own" mean in measurable terms?

Inclusion criteria has been changed to - Participants should have an academic research paper that they have read.

Screening questionnaire (Previously Appendix C) has been removed as all Graduate students, Postdocs and Research Assistants in Computer Science read academic research papers as part of their courses or Thesis or Research group meetings. "Collection of research documents of their own" means that an academic research paper that participants have read. That academic research paper can be participant's own published paper or a published paper that they have studied/read as part of their thesis or courses or just for interest.

In summary, there needs to be greater clarity on (1) the recruitment process and instruments, (2) the inclusion and exclusion criteria (will English comprehension (reading and writing) be important?), and (3) the screening tool.

(1) the recruitment process and instruments and (2) the inclusion and exclusion criteria have already been mentioned above.

English comprehension (reading and writing): English proficiency to be accepted to Computer Science Grad school is sufficient for this study.

(3) the screening tool: Removed as it is not required.

2.4 INFORMED CONSENT PROCESS
2.4.1 It is unclear if the screening questionnaire (Appendix C) is for recruitment purposes or for study related data. If participants will be consented prior to knowing if they should be included, then how are ineligible participants' data being used for the project?

Screening questionnaire (previously Appendix C) was only for recruitment purposes. But since all Graduate students, Postdocs and Research Assistants in Computer Science read academic research papers as part of their courses or Thesis or Research group meetings, hence it has been removed. A link to consent form has been added in the Email and Paper Poster Recruitment. Hence participants can read

Please indicate when participants will be providing consent, as recruitment procedures are not mentioned, it is unclear how participants will be contacting the PI to participate, and if this is the time when they will be consented and begin the study. Will consenting be taking place face-to-face? Will participants have a copy of the consent form prior to signing consent and starting the study? How will you ensure participants are informed? Will you be going through the consent form with them?

Section – 2.4.1 and 2.3.2
Page Numbers – 7 and 8

The participants will be presented with the informed consent form at the beginning of the study. They need to provide consent in the beginning of the study before they proceed with the introduction and training. The consent will be provided by choosing "Yes" from the radio button "Yes" or "No" through the user interface of our system. The link to consent form will also be included as part of the recruitment email and poster. This makes it possible for participants to read the consent form prior to planning to be a part of the study. The consent form contains a brief introduction of the study, information about the confidentiality and anonymity of the participant's data, the participant's right to withdraw and the compensation. The online form will be provided and administered by PI at the very beginning of the study. This will happen face-to-face.

Recruitment procedures has been briefly mentioned in "Recruitment Plans" in section 2.3.2 now. This plan also mentions that participants can send email to the PI on the email mentioned in the Recruitment email and Paper Poster.

There is no information about if participants will be providing written or oral consent, and how the PI will be collecting this.

Section – 2.4.1
Page Number - 8

As mentioned above that this consent form will be online and hence particpants will be providing their consent online and it will be recorded as "Yes" through our system itself. Once they choose "Yes" then only they can proceed with the study further.

2.5 METHODS AND ANALYSIS
2.5.1 There are several pieces of information here that should be expanded on in sections 2.3 and 2.4. For example, it is mentioned that "there are several research

groups in the Faculty of Computer Science which are interested in participating in this user study". This should be outlined in the recruitment section. Will you be recruiting through "MALNIS"? And if so, who are they? How will interested participants from these groups be identified? Who will be recruiting them and involved in disseminating information about the research project?

Section – 2.3.2 and 2.5.1
Page Numbers – 7, 8 and 9

Section 2.3 and 2.4 has been expanded briefly.

Reference to any research groups has been removed as the recruitment will be done through recruitment email to csgrads@cs.dal.ca and paper poster only which will include all Graduate students, Postdocs and Research Assistants in Computer Science at Dalhousie.

A task for the project is having the participant upload a piece of writing that is either theirs or one that they have read. Please outline the requirements of this task and how you will ensure that it fits within the confines of the research topic. If participants upload their own writing, does it have to be peer reviewed? Please identify more about what qualifies and acceptable piece of work for this task.

Section – 2.3.1
Page Number - 7

As mentioned in Section 2.3.1 participant need to upload an academic research paper and not just any piece of writing. That research paper can be their own published paper or a published paper that they have studied as part of their thesis or courses or just for interest. All published papers are already peer reviewed and are verified before publishing.

The timeline is alarming and suggests the PI has not thought about the tasks and requirements from participants. The expected time and allowed 60 minutes are not equal. Please consider how only 3 minutes is not enough time to ensure your participant is fully informed about your study. Even if the participant had access to the consent form prior to consenting, it would take more than 3 minutes to ensure that they had no questions, and were fully aware of the study expectations, their rights, risks, and ability to withdraw. Will 2 minutes be enough time for participants to read the questions and answer?

Please reconsider how much time it will take to perform all these tasks, including the consenting and screening questionnaire, and then provide an accurate estimation.

Section – 2.5.1

Page Numbers – 9 and 10

Expected time for consent form has been increased to 10 minutes, which is sufficient for participants to read and understand it thoroughly. Moreover, a link has been added in the recruitment email and paper poster making it accessible for participants to read and understand it prior to showing interest for the study. Screening and demographic questionnaire has been removed. Total expected time for the study is now 65 minutes. Timeline has been changed accordingly. However, if participants want to spend more time then they are free to do so.

It is unclear if users can complete this on a computer of their own, or if the PI will be providing research space. Please clarify this, and how this may influence how participants are consented and screened for eligibility, and how they receive their compensation for participating.

Section – 2.5.1 and 2.5.3
Page Numbers – 9, 10 and 11

We will provide a quiet room in the Computer Science Faculty for the study. However, participants have to bring their own laptop. The complete study is online through a browser and hence participants do not need to install anything. This has been specified in the consent form as well.

Participants will be given $15 CAD independent of time taken by the them to complete the study. . This is clearly outlined in the consent form. Participant will be paid the full compensation even if they don't complete the study and plan to leave.

Please outline the tests for the components in this section, what they will be measuring, how they are constructed, their reliability and validity, etc.

Section – 2.5.2
Page Numbers – 10 and 11

Complete information on what data will be measured during tests for various components of the study, has been mentioned in detail in section 2.5.2. These tests have been constructed in a way to collect data for proving our hypotheses mentioned in section 2.2. Moreover, testing a software/tool is an important part of the software development life cycle and will help in improving the tool further.

2.5.3 Please provide clarity as there is conflicting information in this section. Will participants be provided with compensation if they do not complete the entire study? Or will they need to complete the "30 minutes introduction and completing questionnaire and signing forms plus the expected 30 minutes interaction with the system". Please be aware of the TCPS section 3.1 that states:

> *The participant should not suffer any disadvantage or reprisal for withdrawing nor should any payment due prior to the point of withdrawal be withheld. If the research project used a lump-sum incentive for participation, the participant is entitled to the entire amount. If a payment schedule is used, participants shall be paid in proportion to their participation.*

Clarity is requested on this point.

Section – 2.5.3
Page Number - 11

Participants will be given $15 CAD independent of time taken by the them to complete the study. . This is clearly outlined in the consent form as well. Participant will be paid the full compensation even if they don't complete the study and plan to withdraw. Participants do not have to incur any expenses to participate.

2.5.5 Please outline the PI's qualifications relevant to the study.

Section – 2.5.5
Page Number - 11

Primary investigator Deepak Munjal will be responsible for recruitment, administering the study and analysis of the results. Deepak will be the primary researcher responsible for reviewing and accessing results. Deepak started working towards Master of Computer Science degree from Fall 2016 and his area of interests are Machine Learning and Visualization. As part of the degree, he has also taken many courses relevant to these areas at Dalhousie University. Before coming to Dalhousie, Deepak has worked on multiple Data Science projects at Google for 4 years.

2.6 PRIVACY AND CONFIDENTIALITY
2.6.1 It is unclear if other participants will have access to documents that are uploaded by a participant. If this is the case, how will you ensure privacy and confidentiality of that information and intellectual property?

Section – 2.6.1
Page Number - 12

There won't be any personal or sensitive data collected in this study. The data will be collected while users rate the quality of classified topics and interact with the user interface of the visualization system and filling questionnaires. Individual research papers will not be recorded. The uploaded research paper can be deleted immediately after the classification process is done by participants. DOI (Digital Object Identifier) of the research paper that participants bring with them, the

classified topics extracted by the system, and the edit made to the topics by the "expert" (participant) will be saved by our system, which will only be accessible by PI. The participants will be identified by a random ID instead of their names. **No participant can see or access documents by other participants as no documents (academic paper) are being stored by us.** A participant will have the option to download the list of classified topics of their own research paper only (academic research paper that they brought with them for the study). If they choose to download it, it will be saved on their own laptop. The records will be deleted from PI's hard drive five years after the experiment or publication. Two copies on CD will be stored in a locked room (that only the supervisor has access to). Those CDs will be destroyed in a shredding machine five years after the experiment or final publication.

While maybe not sensitive, there is personal data being collected in the study. Please indicate how consent forms with participant information will be stored and how the database linking participant names to IDs will be stored. Personal information on the demographic survey in a sample that is small enough (n=10) could potentially identify participants. Please consider the implications of this on privacy, confidentiality, and anonymity – both in the PI being able to identify participants during analysis, and any data that might be disseminated to the public.

Section – 2.6.2
Page Number - 12

Consent form has been edited. There won't be any personal or sensitive data collected in this study. There will be no linking between participant names to IDs as participants' name is not being stored. We only store participants' IDs. Demographic survey has been removed. Since the study population is less (10 participants), identification to the level of the research group/interest might be possible, but it is not possible to identify an individual due to the coarse-grained nature of the topics (40 topics for all of Computer Science). Participants will be quoted in final report, but only by the random ID assigned to them. Additional permission is being requested on the consent form.

It is also not clear what information participants will be able to download from the study itself. Is this information that other participants have uploaded? How does this information relate to the data being collected for the study?

Section – 2.6.1 and 2.5.1
Page Numbers – 9, 10 and 12

A participant will have the option to download the list of classified topics of their own research paper only (academic research paper that they bring with them for the study). This data is not the result of the study. This data is just a list of classified

2.6.2 Why is an ID used to link participants to their names/results? This does not seem necessary, and the Board recommends not keeping a linking key.

Section – 2.6.2
Page Number - 12

Each participant will have only an ID. There will be no linking between names and ID, as names or any other personal information is not saved.

2.7 PROVISION OF RESULTS TO PARTICIPANTS
2.7.1 It is mentioned that participants will be able to download data via the interface, but it is unclear what these data are, and if these data are the results of the study. Please clarify. Additionally, it is mentioned that a publication will be made available to participants on request. Identify how participants will know they can request this and how they might do so.

Section – 2.7.1
Page Number - 13

A participant will have the option to download the list of classified topics of their own research paper only (academic research paper that they brought with them for the study). This data is not the result of the study. This data is just a list of classified topics of the participant's research paper by the Machine Learning system. Participants will see the announcement of the Thesis defense and hence can see the study results by attending the defense. If they are interested in knowing more about the study results, then they can discuss this during Thesis defense.

2.8 RISK AND BENEFIT ANALYSIS
2.8.1 It is unclear what data the participants will be uploading to the user interface. If this is their own personal work and intellectual property, please identify the risks that might be associated with that.

Section – 2.8.1
Page Number - 14

Participants will be uploading an academic research paper that they have read. These research papers are published papers and hence no risks are involved.

2.9 CONFLICT OF INTEREST
It is unclear what the statement "Possible students with the same supervisor" means and how it relates to the project, conflict of interest and dual roles. Given more clarity on the other sections, this may make more sense, but as the proposal is written now, the relationship between students and supervisors have not been identified, nor has it been indicated that students and supervisors will know who's participating. Please clarify.

Please indicate ACMDSP's role in this study.

Please indicate if the PI has any teaching responsibilities (teaching assistant, instructor) and if that role may be a conflict of interest in recruiting participants.

Section – 2.9
Page Number - 14

The previous statement in the conflict of interest - "Possible students with the same supervisor" has been removed. No identifying information is being stored and there is no conflict of interest between anyone. ACMDSP is a potential user of the system. PI has no teaching responsibilities.

RECRUITMENT MATERIAL
Please append all recruitment materials. This is an email (and its use in described in 2.3.2), but the researcher also mentions recruiting through Notice Digest (now Today@Dal) – please append this notice.

Section – 2.3.2
Appendices – A and B
Page Numbers – 7, 8, 19 and 20

Recruitment through Notice Digest has been removed. Recruitment will be done by sending email to csgrads@cs.dal.ca and by posting Recruitment paper posters on notice boards in the Goldberg Computer Science building. All recruitment materials are appended in appendices.

There are a number of grammatical errors throughout that make it difficult to know the purpose of the study, the inclusion/exclusion criteria, etc. Please do a thorough proof read and edit of recruitment materials.

Major editing has been done throughout the application form and recruitment materials. It has been proofread and corrected for all grammatical errors as well.

SCREENING MATERIAL
Appendix C

Please consider how screening information could be providing you with the information you need. For example, there is no information on experience with user interfaces, awareness of the themes you're looking to use in the study, experience with the field of study, etc.

Screening questionnaire (previously Appendix C) has been removed.

CONSENT FORM
Appendix B
Under "Study Design" please outline what participants will be doing if they agree to participate.

Under "Who can participate in the study", there is additional exclusion criteria here that was not outlined in the application, please make consistent throughout.

Please clearly indicate that participants can withdraw at any time, and that they will still be compensated.

Why are only certain tasks bolded under "What you will be asked to do". You also have signing the consent form twice, once at the start and once at the end.

Please indicate to participants, under confidentiality, how their anonymity will be kept if results are published. Please also indicate that given the sample size, participants may be identified.

Indicate any conflicts of interest.

Appendix - C
Page Number – 21, 22, 23, 24 and 25

Consent form has been edited accordingly to include the recommended revisions and clarifications for each of the above point.

RESEARCH INSTRUMENTS
Appendix E: It is mentioned that these are examples of some of the questions that will be used on the Visualization Component. Please include a copy in full of the test that participants will be using.

Questions have been edited. The statement that "these are examples" has been removed since these are the exact questions which will be used for the study.

Visual Analytics of Research Community Expertise in Space and Time

Principal Investigator:

Deepak Munjal

Supervisor:

Dr. Evangelos Milios

Faculty of Computer Science

January 2019

Table of Contents:

RESEARCH ETHICS BOARDS APPLICATION FORM

Prospective Research

This form should only be used if new data will be collected. For research involving only secondary use of existing information (such as health records, student records, survey data or biological materials), use the *REB Application Form – Secondary Use of Information for Research.*

This form should be completed using the *Guidance for Submitting an Application for Research Ethics Review* available on the [Research Ethics website](#) (application instructions).

SECTION 1. ADMINISTRATIVE INFORMATION   [File No: office only]

Indicate the preferred Research Ethics Board to review this research:

[ ] Health Sciences OR [✓] Social Sciences and Humanities

Project Title:
Visual Analytics of Research Community Expertise in Space and Time

1.1 Research team information			
Dalhousie researcher name	**Deepak Munjal**		
Banner #	B00748375	Department	Faculty of Computer Science
Email (@dal)	deepak.munjal@dal.ca	Phone	902-818-3840
Study start date	Jan 28, 2019	Study end date	Feb 01, 2019
Co-investigator names and affiliations			
Contact person for this submission (if not lead researcher)	Name		
	Email		Phone	

1.2 For student submissions:

Degree program	Master of Computer Science (MCS)		
Supervisor name and department	Evangelos Milios Faculty of Computer Science		
Supervisor Email (@dal)	eem@cs.dal.ca	Phone	902-494-7111

Department/unit ethics review (if applicable). **Undergraduate minimal risk research only**.
Attestation: [ ] I am responsible for the unit-level research ethics review of this project and it has been approved. Authorizing name: Date:

1.3 Other reviews: Not Applicable		
Other ethics reviews (if any)	Where	Status
Funding, if any (list on consent form)	Agency	
	Award Number	
Peer review (if any)		

1.4 Attestation(s). The appropriate boxes *must* be checked for the submission to be accepted by the REB)

[✓] I am the **lead researcher**. I agree to conduct this research following the principles of the Tri-Council Policy Statement *Ethical Conduct for Research Involving Humans* (TCPS) and consistent with the University *Policy on the Ethical Conduct of Research Involving Humans*.

I have completed the TCPS Course on Research Ethics (CORE) online tutorial.

[✓] Yes   [ ] No

For Supervisors (of student / learner research projects):

[✓] I am the **supervisor** for this research named in section 1.2. I have reviewed this submission, including the scholarly merit of the research, and believe it is sound and appropriate. I take responsibility for ensuring this research is conducted following the principles of the TCPS and University Policy.

I have completed the TCPS Course on Research Ethics (CORE) online tutorial.

[✓] Yes   [ ] No

SECTION 2. PROJECT DESCRIPTION

2.1 Lay summary

2.1.1 In lay language, describe the rationale, purpose, study population and methods. Include the background information or literature to contextualize the study. Mention what new knowledge is anticipated, and whether this is a pilot project or fully developed study. [500 words]

- Text classification is an important task of classifying text into one or more classes or categories or topics. ACMDSP is the ACM Distinguished Speaker Program, which has a set of speakers to provide lectures at various locations worldwide. Currently, the lectures are classified manually into various topics. This research will allow the automated classification of lectures into various topics using the title and abstract information of a lecture. Machine Learning is used to train a classifier which will perform this job. Currently, ACMDSP data is being handled through Excel sheets. This research will allow more intuitive interaction with the database using a Visualization system.

- The research study will have two sub-parts. The first sub-part of the study deals with the Machine Learning system. ~~in first~~ In this, users will be asked to upload ~~their~~ an academic research paper~~s or documents~~. It can be done ~~by~~ through the user interface within the tool itself. The title and abstract from these papers ~~or documents~~ will be passed over to the trained classifier/Machine Learning system. The classifier/system will categorize it into one or more classes/topics. These classified topics will be chosen from a set of 40 topics that we have defined for our Machine Learning System. This set of classified topics will be shown to the user in the user interface.~~form of a document.~~ The user will be asked to evaluate these topics, ~~and~~ make edits to the topics (add/remove any topic) and enter the number of correct topics in the user interface itself. This will help in testing the accuracy of the classifier/Machine Learning system from a user point of view.

- The second sub-part of the study deals with the Visualization system. These two subparts are in continuation and user can do it in a single flow. A connection to the ACMDSP database is provided with a UI (User Interface). The users have to interact with the various options provided within the UI to query the database.

- Our estimated study population will be 10 persons. ~~with expertise in one of the domains of the documents (research papers) they read and have basic knowledge of using a computer. They will be Dalhousie University students, staff, or professors.~~ The population will be Graduate students, Postdocs and Research Assistants in Computer Science at Dalhousie University who have read an academic research paper. That research paper can be their own published paper or a published paper that they have studied as part of their thesis or courses or just for interest. After completing the study, participants will fill out a questionnaire to further express their views of the tool. This user study is very important given the fact that testing a tool is an important part of a software development life cycle and will provide useful feedback for future improvements.

- The system designed as a part of the research automatically classifies a research ~~paper work~~ into topics/classes using Machine Learning. But the automatic approaches are not always correct. Similarly, for the Visualization system, there's a need for the user evaluation which will help in gathering feedback and improving the user interface further. Hence we need user involvement to evaluate these systems, which can be done through a user study. These two components – Machine Learning and Visualization, are part of a single system together as an online tool which will be accessible through a user interface.

- This study will help to get the user feedback about the research. The outcome of the study will reflect on how good the systems are as per user evaluation. This feedback will be very useful in improving the Machine Learning and Visualization System further.

- This is intended to be a fully developed project that can be used to categorize research expertise and visualize it.

- Reference:

1. Van Dijk, D., Tsagkias, M., & de Rijke, M. (2015, August). Early detection of topical expertise in community question answering. In Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 995-998). ACM.

2. Ziaimatin, H., Groza, T., Bordea, G., Buitelaar, P., & Hunter, J. (2018). Expertise Profiling in Evolving Knowledge curation Platforms. GSTF Journal on Computing (JoC), 2(3).

3. Hsiao, I. H., & Awasthi, P. (2015, March). Topic facet modeling: semantic visual analytics for online discussion forums. In Proceedings of the Fifth International Conference on Learning Analytics And Knowledge (pp. 231-235). ACM.

4. Kuhn, K. D. (2018). Using structural topic modeling to identify latent topics and trends in aviation incident reports. Transportation Research Part C: Emerging Technologies, 87, 105-122.

5. Giaquinto, R., & Banerjee, A. (2018). Topic Modeling on Health Journals with Regularized Variational Inference. arXiv preprint arXiv:1801.04958.

2.1.2 If a phased review is being requested, describe why this is appropriate for this study, and which phase(s) are included for approval in this application.

[✓] Not applicable

2.2 Research question

State the hypotheses, the research questions or research objectives.

- We have one hypothesis and one research objective:
  - **Hypothesis**: It is faster and more accurate to interact with a database with our tool than with the Excel Sheets.

> - The user's input will take two forms:
>   - It will include the user's interaction with the Visualization system. This will include clicking on a button, entering a search query in the search box, etc.
>   - It will also include the research paper/document, provided by the user, to pass its title and abstract through the Machine Learning System/Classifier. This will also be done directly using the user interface.

2.3 Recruitment

2.3.1 Identify the study population. Describe how many participants are needed and how this was determined.

- The population for our study will be ~~Dalhousie University students, staff, and professors especially those who are going to write a related work section in their reports/thesis or writing a survey paper~~ Graduate students, Postdocs and Research Assistants in Computer Science at Dalhousie University who have read an academic research paper. That research paper can be their own published paper or a published paper that they have studied as part of their thesis or courses or just for interest. We will need 10 participants to have a credible statistical analysis. This number of participants is selected based on similar user studies performed in this area.

2.3.2 Describe recruitment plans and append recruitment instruments. Describe who will be doing the recruitment and what actions they will take, including any screening procedures. Describe and justify any inclusion / exclusion criteria.

~~Screening: After the participants signed the consent form, screening will be done through the Screening Questionnaires. The participants should be in computer science or similar majors who are familiar with the concept of classification and read scientific papers. However, there is no criterion based on the screening questionnaire and we will give an introduction about the process before they start. The screening questionnaire is needed to know how familiar our participants are with the text document classification in general.~~

- ~~Emails will be sent to students who subscribe to the computer science mail list (csall@dal.ca) and also through Dalhousie events calendars (i.e. Notice Digest, notice.digest@dal.ca).~~

- ~~The~~ Recruitment Plans: PI will send an e-mail to all Graduate students, Postdocs and Research Assistants in Computer Science at csgrads@cs.dal.ca. Recruitment paper posters will also be posted on notice boards in the Goldberg Computer Science building. Interested candidates can read the email or posters, read consent form by opening the link available in both recruitment email and recruitment poster. If they are interested, they can email to PI on the email mentioned in both email and poster. They need to send their Degree name and year of study in the email. ~~the Faculty of Computer Science and Dalhousie events calendars.~~ If we receive emails from more than 10 candidates who are interested in taking part in the study, then we will select only 10 participants out of those candidates based on the seniority level (Degree name and Year of study). Further, PI will reply to them and invite them for the study at a date and time based on their availability.

- Recruitment Instruments/Tools: PI will send an email at csgrads@cs.dal.ca and paper posters will be used as well. These paper posters will be put on the notice boards in Goldberg Computer Science building, Dalhousie University. The email and poster are attached in appendices.

- Inclusion criteria: Participants should have an academic research paper that they have read. ~~Participants should have a collection of research documents of their own or they read/follow: students writing a thesis, professors writing a proposal or planning a reading course or organizing a conference. They should be familiar with their research document because they are required to judge the quality of the classified topics by the Machine Learning system/classifier. They need to know what the topics mean and whether they are describing the same topic.~~

2.3.3 Describe any community or organizational permissions needed to recruit your participants (attach support letters). Describe any other community consent or support needed to conduct this research. (If the research involves Aboriginal participants, please complete section 2.10).

[✓] Not applicable

2.4 Informed consent process

2.4.1 Describe the informed consent process, including any plans for ongoing consent (how and when the research will be described to prospective participants, by whom, how the researcher will ensure prospective participants are fully informed). If non-written consent is proposed, describe the process. Address how any third party consent (with or without assent) will be managed. Append copies of all consent/assent documents, including oral consent scripts.

- Please see appendices. The participants will be presented with the informed consent form at the beginning of the study. They need to provide consent at the beginning of the study before they proceed with the introduction and training. The consent will be provided by choosing "Yes" from the radio button "Yes" or "No" through the user interface of our system. ~~complete the screening questionnaire.~~ The link to the consent form will also be included as part of the recruitment email and poster. This makes it possible for participants to read the consent form prior to planning to be a part of the study. ~~This~~ The consent form contains a brief introduction of the study, information about the confidentiality and anonymity of the participant's data, the participant's right to withdraw and the compensation. The online form will be provided and administered by the PI at the very beginning of the study.

2.4.2 Discuss how participants will be given the opportunity to withdraw (their participation and/or their data) and any limitations on this.

- The participants will be told in the beginning that they can withdraw from the study at any time they want without penalty. This is mentioned in the consent form.

2.4.3 If an exception to the requirement to seek prior informed consent is sought, address the criteria in TCPS article 3.7A.

[✓] Not applicable

2.5 Methods and analysis

2.5.1 Describe the study design, where the research will be conducted, what participants will be asked to do and the time commitment, what data will be recorded using what research instruments (append copies).

1. Study Design

   There are two tasks as part of the user study:

   Users have to provide consent by choosing the "Yes" radio button in our system's user interface and further will be provided with introduction and training.

   Users have to fill a pre-screening questionnaire, which will help researchers understand their backgrounds and to find if they are suitable for the user study.

   i. Visualization System testing:

   - Since the system is designed to help ACMDSP committee with searching for a speaker with their expertise or location. Users are expected to interact with the system in the same way. Users will use both traditional Excel sheets and our tool to interact with the data and to answer our questions. They will be provided with a set of questions to answer, which will make the user interact with the system and will help researchers to test the easiness of the Visualization system and to know user feedback on the same.

   ii. Machine Learning testing:

   - During the study, each user has to upload a digital copy of an academic research paper/document of their field of expertise. Title and abstract of this research paper will be passed to the Machine Learning system as an input. The system will predict a set of topics. The user has to evaluate the predicted topics and, make edits to the topics (add/remove any topic) and write the number of correct predicted topics, which users have to enter within the user interface.

   - Finally, we ask them to fill out a questionnaire and give comments individually. There are several research groups in the Faculty of Computer Science which are interested in participating in this user study, like the MALNIS research group (https://uniweb.dal.ca/groups/5/profile).

2. We will provide a quiet room in the Computer Science Faculty for the study. However, participants have to bring their own laptop. The complete study is online through a browser and hence participants do not need to install anything.

3. Participants will be asked to perform the following tasks in 65 60 minutes:

- **Participants will** sign read **the consent form** in our system's user interface and provide consent by choosing "Yes" radio button. **(**3 10 **minutes)**

- Participants will complete screening and demographic questionnaires. (2 minutes)

- After providing consent, P **participants will be given several short examples to become familiar with the user interface and the whole system. (10 minutes)**

- Each participant will be given one user ID, which will be used for testing both Machine Learning and Visualization system. The post-condition (evaluation) questionnaire will be given to the users as well. Before the visualization system test, users will be provided with a set of questions that will make users interact with the system.

- The participants are asked to perform document classification on the academic research paper that they bring with them. their own documents (research papers). Then they have to evaluate the classified topics, make edits to the topics (add/remove any topic) and write the number of correct topics in the user interface of our system itself. It is expected that the participants spend about 10 minutes for this step. Similarly, for the Visualization system, users are expected to spend about 20 minutes. However, the system is available for any participant who likes to spend more time.

- After finishing the user study, the participant will fill the evaluation questionnaire, submit it and get the compensation. This is expected to take 15 minutes.

- Hence, Signing forms providing consent, introduction, and training will take 15 20 minutes. Filling out the evaluation questionnaire will also take around 15 minutes. The system testing for both Visualization system and Machine Learning system is expected to be done in 30 minutes; however, participants would do it on their own time so they can spend more time with the system/interface if they like.

4. A log of participant's operations time for interacting with the Excel sheet and time for interacting with in the user interface of the Visualization system and along with DOI (Digital Object Identifier) of the research paper that participants bring with them, the classified topics extracted by the system, and the edit made to the topics by the "expert" (participant) and the number of correct topics, as evaluated by the participant/user, for the Machine Learning system will be saved. Only a computer connected to the Internet is needed. This data will be collected by the online system itself. Participants need to bring their laptop and it needs to be connected with the Internet. No other data like participant's personal information will be asked or saved.

[ ] This is a clinical trial (physical or mental health intervention) – ensure section 2.11 is completed

2.5.2 Describe plans for data analyses.

- For Machine Learning system, data analysis will focus on how many predicted topics users think are correct. This number will be collected and will help in providing an estimate of the accuracy of the system. Moreover, DOI (Digital Object Identifier) of the research paper that participants bring with them, the classified topics extracted by the system, and the edit made

to the topics by the "expert" (participant) will be very useful in improving the Machine Learning system and further testing it after improvement. For visualization system, time that users have interacted with the Excel Sheets and time they have used to interact with the user interface of the Visualization system will be collected. This data will be analyzed to see the total time required and to access the easiness of the user interface of the Visualization System over the Excel sheet. ~~user interaction data and time to answer visualization test questions will be collected. This will include searching query through search boxes, clicking on buttons etc. The basis of this interaction will be the set of questions, provided by researchers, that will make users interact with the Visualization system. This will be analyzed to see the number of actions taken and the total time required. This will be help in understanding user behavior towards our system.~~

2.5.3 Describe any compensation that will be given to participants and how this will be handled for participants who do not complete the study. Discuss any expenses participants are likely to incur and whether/how these will be reimbursed.

- Participants will be given $15 CAD independent of time taken by them to complete the study. ~~(for 30 minutes introduction and completing questionnaire and signing forms plus the expected 30 minutes interaction with the system).~~ This will be clearly outlined in the consent form. ~~The compensation will be given when the participant submits the evaluation questionnaire.~~ Participants will be paid the full compensation even if they don't complete the study and plan to withdraw. ~~plan to leave and doesn't complete the study.~~

- Participants do not have to incur any expenses to participate.

2.5.4 Describe and justify any use of deception or nondisclosure and explain how participants will be debriefed.

[✓] Not applicable

2.5.5 Describe the role and duties of local researchers (including students and supervisors) in relation to the overall study. Identify any special qualifications represented on the team relevant to the proposed study (e.g. professional or clinical expertise, research methods, experience with the study population, statistics expertise, etc.).

- Primary investigator Deepak Munjal will be responsible for recruitment, administering the study and analysis of the results. Deepak will be the primary researcher responsible for reviewing and accessing results. Deepak started working towards Master of Computer Science degree from Fall 2016 and his area of interests are Machine Learning and Visualization. As part of the degree, he has also taken many courses relevant to these areas at Dalhousie University. Before coming to Dalhousie, Deepak has worked on multiple Data Science projects at Google for 4 years. Dr. Evangelos Milios and Dr. Fernando Paulovich will provide guidance and assistance during analysis. Both professors are very experienced with the study population and the domain of Machine Learning and Visualization.

2.6 Privacy & confidentiality

2.6.1 Describe any provisions for ensuring privacy and confidentiality (or anonymity). Describe who will have access to data and why, how data will be stored and handled in a secure manner, how long data will be retained and where. Discuss any plans for data destruction and/or de-identification.

[ ] This research involves personal health records (ensure section 2.12 is completed)

1. There won't be any personal or sensitive data collected in this study.

2. The data will be collected while users rate the quality of classified topics and interact with the user interface of the visualization system and filling questionnaires. Individual ~~document collections~~ research papers will not be recorded. The uploaded ~~personal documents~~ research paper can be deleted immediately after the classification process is done by participants. DOI (Digital Object Identifier) of the research paper that participants bring with them, the classified topics extracted by the system, and the edit made to the topics by the "expert" (participant) will be saved by our system, which will only be accessible by PI. The participants will be identified by a random ID instead of their names. No participant can see or access documents by other participants as no documents (academic paper) are being stored by us. A participant will have the option to download the list of classified topics of their own research paper only (academic research paper that they bring with them for the study). If they choose to download it, it will be saved on their own laptop.

3. The records will be deleted from PI's hard drive five years after the experiment or publication. Two copies on CD will be stored in a locked room (that only the supervisor has access to). Those CDs will be destroyed in a shredding machine five years after the experiment or final publication.

2.6.2 Describe how participant confidentiality will be protected when research results are shared. Discuss whether participants will be identified (by name or indirectly). If participants will be quoted address consent for this, including whether quotes will be identifiable or attributed.

- Participants' names will not be used in any publication. ~~The unique ID is used to identify a profile with the participant's name.~~ Each participant will have only an ID. There will be no linking between names and ID, as names or any other personal information is not saved.

- Since the study population is less (10 participants), identification to the level of the research group/interest might be possible, but it is not possible to identify an individual due to the coarse-grained nature of the topics (40 topics for all of Computer Science).

- Participants will be quoted in the final report, but only by the random ID assigned to them. Additional permission will be requested on the consent form. ~~The quotes will be attributed to the random ID assigned to the participants. Permission will be requested on the consent form.~~

2.6.3 Address any limits on confidentiality, such as a duty to disclose abuse or neglect of a child or adult in need of protection, and how these will be handled. Detail any such limits in consent documents.

[✓] Not applicable

2.6.4 Will any information that may reasonably be expected to identify an individual (alone or in combination with other available information) be accessible outside Canada? This includes sharing information with team members, collecting data outside Canada, use of survey companies, use of software.

[✓] No

[ ] Yes. If yes, describe how you comply with the University *Policy for the Protection of Personal Information from Access Outside Canada*, such as securing participant consent and/or securing approval from the Vice President Research.

2.7 Provision of results to participants

2.7.1 The TCPS encourages researchers to share study results with participants in appropriate formats. If you plan to share study results with participants, discuss the process and format.

[ ] Not applicable

1. ~~The participants are able to download the results of their classified topics during the study itself.~~ A participant will have the option to download the list of classified topics of their own research paper only (academic research paper that they brought with them for the study). This data is not the result of the study. This data is just a list of classified topics of the participant's research paper by the Machine Learning system. ~~The publication resulting from the study will be made available to participants on request.~~ Participants will see the announcement of the Thesis defense and hence can see the study results by attending the defense. If they are interested in knowing more about the study results, then they can discuss this during Thesis defense.

~~The user interface will generate the output classified topics in a document file and it is accessible for download inside the user interface. The participants can delete their own classified topics at the end if they like.~~

2.7.2 If applicable, describe how participants will be informed of any incidental findings – unanticipated results (of screening or data collection) that have implications for participant welfare (health, psychological or social).

[✓] Not applicable

2.8 Risk & benefit analysis

2.8.1 Discuss what risks or discomforts are anticipated for participants, how likely risks are and how risks will be mitigated. Address any particular ethical vulnerability of your study population. If applicable, address third party or community risk. Risks to privacy from use of identifying information should be addressed.

- They might get a little bit tired. Otherwise no more than the risk of every day's life.
- Participants will be uploading any academic research paper that they have read. These research papers are published papers. It is not possible to identify an individual due to the coarse-grained nature of the topics (40 topics for all of Computer Science). There is no risk to anyone's privacy.

2.8.2 Identify any direct benefits of participation to participants (other than compensation), and any indirect benefits of the study (e.g. contribution to new knowledge.

- The Machine Learning system might predict topics which are relevant to the research paper and which users were unaware of. This study will provide them knowledge about new areas of expertise that are relevant to their own area of expertise.

2.9 Conflict of interest

Describe whether any dual role or conflict of interest exists for any member of the research team in relation to potential study participants (e.g. TA, fellow student, teaching or clinical relationship), and/or study sponsors, and how this will be handled.

[ ] Not applicable

- No identifying information is being stored and there is no conflict of interest between anyone. ~~Possible students with the same supervisor.~~
- ACMDSP is a potential user of the system.
- PI has no teaching responsibilities.

2.10 Research with Aboriginal peoples

[✓] Not applicable – go to 2.11

2.10.1 If the proposed research involves Aboriginal peoples, describe the plan for community engagement (per TCPS Articles 9.1 and 9.2). Attach supporting letters, research agreements and other relevant documents, if available. If community engagement is not sought, explain why the research does not require it, referencing article 9.2.

2.10.2 State whether ethical approval has been or will be sought from Mi'kmaw Ethics Watch or other Indigenous ethics review group(s), and if not, why the research does not fall under their purview.

2.10.3 Describe any plans for returning results to the community and any intellectual property rights agreements negotiated with the community, with regard to data ownership. If there are specific risks to the community involved, ensure these have been addressed in section 2.8.1.

2.11 Clinical trials

[✓] Not applicable – go to 2.12

2.11.1 Does the proposed research require clinical trial registration, in keeping with national and international regulations?

[ ] No. Please explain why not.

[ ] Yes. Please indicate where it was registered and provide the registration number.

2.11.2 If a novel intervention or treatment is being examined, describe standard treatment or intervention, to indicate a situation of clinical equipoise exists (TCPS Chapter 11). If placebo is used with a control group rather than standard treatment, please justify.

2.11.3 Clearly identify the known effects of any product or device under investigation, approved uses, safety information and possible contraindications. Indicate how the proposed study use differs from approved uses.

[ ] Not applicable

2.11.4 Discuss any plans for blinding/randomization.

2.11.5 What plans are in place for safety monitoring and reporting of new information to participants, the REB, other team members, sponsors, and the clinical trial registry? These should address plans for removing participants for safety reasons, and early stopping/unblinding/amendment of the trial. What risks may arise for participants through early trial closure, and how will these be addressed? Are there any options for continued access to interventions shown to be beneficial?

2.12 Use of personal health information

[✓] Not applicable

2.12.1 Describe the personal health information required and the information sources, and explain why the research cannot reasonably be accomplished without the use of that information. Describe how the personal health information will be used, and in the most de-identified form possible.

2.12.2 Will personal health information be combined with information from other sources to form a composite record (data linkage)? Will the research create individually identifying health information by combining information from two or more databases without the consent of the individuals who are the subjects of the information (data matching)?

[ ] No.

[ ] Yes. Describe the other information and how linkage will be conducted, and/or why data matching is required.

2.12.3 Describe reasonably foreseeable risks to privacy and how these will be mitigated.

SECTION 3. APPENDICES

3.1 Appendices Checklist. Append all relevant material to this application. This may include:

[✓] Recruitment documents (posters, oral scripts, online postings, invitations to participate, etc.)

[✓] Screening documents

[✓] Consent/assent documents or scripts

[✓] Research instruments (questionnaires, interview or focus group questions, etc.)

[ ] Debriefing forms

[ ] Permission letters (Aboriginal Band Council, School Board, Director of a long-term care facility)

[ ] Support letters

3.2 Consent Form

Sample consent forms are provided on the Research Ethics website and may be used in conjunction with the information in the *Guidance* document to help you develop your consent form.

CONSENT FORM CHECKLIST

Please complete this checklist and submit with the application.

YES	N/A	Have you included the following in your consent form/process?
X		Identification of the document as CONSENT FORM
X		Title of study
X		Identity and affiliation of researchers
X		Contact information of individual conducting the study
X		Invitation to participate in research
X		Assurance of voluntariness and right to withdraw without repercussions
X		Short description of the purpose of the study
X		Short description of the study design and how many participants are involved
X		Inclusion and exclusion criteria
X		Description of what the participant is being asked to do
X		Estimate of the participant's time commitment
X		Description of where the research will take place
	X	Description of special clothing or other preparations required of the participant
X		Description of how anonymity will be handled
X		Description of how confidentiality of the data will be assured
X		Description of any necessary limitations of confidentiality protections
X		Description of the nature and probability of risks for participants
X		Description of the benefits for participants
	X	Declaration of any researcher conflict of interest
	X	Description of any possible commercial outcomes of the research
	X	Description of how participants will review transcripts of interviews
X		Description of how study results will be provided to participants

	X	Permissions requested for audio/videotaping
X		Permissions requested for use of quotations
	X	Permission for future use of data in specified studies
	X	Permission to recontact participant for participation in future studies
	X	Permissions related to transportation/use of data outside of Nova Scotia
	X	How assent of participant will be sought when 3rd parties give consent
X		Signature statement indicating that information has been provided
X		Signatures of participant and person obtaining consent

YES	N/A	Have you addressed the following in your Consent Form /Process?
X		Appropriate Reading comprehension level (Grade 8)
X		Avoidance of technical language
X		Formatting: font size (min 12 pt), headings, page numbering
	X	Clear distinction between clinical care/research procedures
	X	No waiver of rights is sought

Appendix A – Email Recruitment Notice

We are recruiting participants to take part in a research study which comprises of evaluation of our system having Visualization and Machine Learning components. We are looking for Graduate students, Postdocs and Research Assistants in Computer Science at Dalhousie University who have read an academic research paper. That research paper can be your own published paper or a published paper that you have studied as part of your thesis or courses or just for interest. ~~students/staffs/professors in computer science or similar majors, who are familiar with concept of text classification in general. Students or professors who are writing thesis/report or those who read research papers/documents frequently and need to classify those documents into categories/topics are preferred.~~ The participant should have a comprehensive of ~~a~~ the academic research paper~~/document, which they own or follow/have read,~~ in PDF format. The study will be conducted online in Computer Science Faculty and ~~will take about an hour~~ is expected to take 65 minutes to complete. You need to bring your own laptop. The whole study will be online through the browser and you do not need to install anything. The link to the consent form is attached to this email. Participants are encouraged to read the consent form and if interested, an email can send to the Principal Investigator mentioning the Degree/Program and the year of study. **If selected for the study,** **You meet with ~~a~~ the Principal Investigator ~~researcher~~ to go over the study details, first give consent to do the study ~~and fill in a background questionnaire~~. You will then have a short training on how to interact with the user interface for about 10 minutes.** We will provide a room in the Computer Science Faculty for the study. After the study is completed, you will fill out and submit an evaluation questionnaire on your experience with the system. You are able to finish the classification process and interaction with Visualization system in ~~60~~ 65 minutes including the introduction session; however, you can spend more time if you like. In the end, you will get a set of classified topics/categories for your research paper~~/document~~. Compensation is $15 CAD for participation in the study. The topics predicted by the Machine Learning system may contain some new areas relevant to your area of expertise. This study is a good opportunity to classify your ~~documents~~ paper and to learn more about those different research areas out there which might be of interest to you.

If you are interested in participating, please contact Principal Investigator Deepak Munjal (deepak.munjal@dal.ca).

Consent Form link: http://bit.do/consent-form-study

Appendix B – Poster Recruitment Notice

PARTICIPANTS NEEDED FOR RESEARCH STUDY IN MACHINE LEARNING AND VISUALIZATION SYSTEM TESTING

We are recruiting participants to take part in a research study which comprises of evaluation of our system having Visualization and Machine Learning components.

We are looking for Graduate students, Postdocs and Research Assistants in Computer Science at Dalhousie University who have read an academic research paper.

The participant should have a comprehensive of the academic research paper in PDF format. The study will be conducted online in Computer Science Faculty and is expected to take **65 minutes** to complete. You need to bring your own laptop.

The whole study will be online through the browser and you do not need to install anything. The link to the consent form is attached to this poster. Participants are encouraged to read the consent form and if interested, an email can send to the Principal Investigator mentioning the Degree/Program and the year of study**.**

If selected for the study, You meet with the Principal Investigator to go over the study details, first give consent to do the study. You will then have a short training on how to interact with the user interface for about 10 minutes.

Compensation is **$15 CAD** for participation in the study. The topics predicted by the Machine Learning system may contain some new areas relevant to your area of expertise. This study is a good opportunity to classify your paper and to learn more about those different research areas out there which might be of interest to you.

If you are interested in participating, please contact Principal Investigator Deepak Munjal (deepak.munjal@dal.ca).

More information on Consent Form link: http://bit.do/consent-form-study

This study has been reviewed by, and received ethics clearance through the Dalhousie University Research Ethics Committee.

Appendix B C – Consent Form

Visual Analytics of Research Community Expertise in Space and Time

Principal investigator: Deepak Munjal {deepak.munjal@dal.ca}(902-818-3840)

Supervisor: Dr. Evangelos Milios (eem@cs.dal.ca) (902-494-7111)

Faculty of Computer Science, Dalhousie University, 6050 University Ave., PO Box 15000, Halifax, NS, B3H 4R2, Canada.

Contact: Deepak Munjal {deepak.munjal@dal.ca}(902-818-3840)

You are invited to participate in a research study being conducted by Deepak Munjal who is a Master of Computer Science student at the Faculty of Computer Science, Dalhousie University. This study is being done as part of the research thesis.

In order to participate in this study, you should be in either be a Graduate Student in computer science or a Postdoc or Research Assistant in Computer Science at Dalhousie University. a similar major who is familiar with the concept of classification and visualization. The information in this consent form will outline any possible risks, inconvenience, and discomfort that you might experience.

Participation is voluntary, and participants are free to withdraw at any time without repercussions. If you have any concern and question about the study, please do not hesitate to ask the principal investigator.

Purpose This study is to evaluate visualization and Machine learning components of our system to determine the quality of classified topics and to know feedback about user interface based on users' point of view.

Study Design The p After providing consent, participants will be given several short examples to become familiar with the user interface and the whole system. Each participant will be given one user ID, which will be used for testing Machine Learning and Visualization system. The post-condition (evaluation) questionnaire will be given to the users as well. Before the visualization system test, users will be provided with a set of questions that will make users interact with the system. The participants are asked to perform document classification on the academic research paper that they bring with them. Then they have to evaluate the classified topics, make edits to the topics (add/remove any topic) and write the number of correct topics in the user interface of our system itself. It is expected that the participants spend about 10 minutes for this step. Similarly, for the Visualization system, users are expected to spend about 20 minutes. However, the system is available for any

participant who likes to spend more time. After finishing the user study, the participant will fill the evaluation questionnaire, submit it and get the compensation. This is expected to take 15 minutes. The Principal investigator will analyze the ~~operations~~ time the participants take for each part of the study~~use during the experiment~~ in order to design a better interface aiding user interactions. This time will be recorded automatically by the system. The principal investigator will also analyze the participant evaluation results of different classified topics for each academic research paper.~~/document.~~

Who can participate in the study You may participate in this study if you are ~~a~~ a Graduate Student or Postdoc or Research Assistant in Computer Science. You should be having an academic research paper that you have read. That research paper can be your own published paper or a published paper that you have studied/read as part of your thesis or courses or just for interest. If they are interested, they can email to PI on the email mentioned in both email and poster. They need to send their Degree name and year of study in the email. If PI receives emails from more than 10 candidates who are interested in taking part in the study, then PI will select only 10 participants out of those candidates based on the seniority level (Degree name and Year of study). Further, PI will reply to them and invite them for the study at a date and time based on their availability. ~~student/staff/faculty in computer science or similar majors who are familiar with the concept of text classification, visualization and read scientific papers monthly. You should be able to understand the categorization of a text document into various topics/categories and should have basic knowledge about using a simple user interface.~~

Who will conduct the research Deepak Munjal, a~~n~~ MCS (Master of Computer Science) candidate from the Faculty of Computer Science, Dalhousie University, will be conducting the research, i.e., ~~distributing and collecting questionnaires~~ Introducing the online consent form to provide consent, introducing and demonstrating the system~~software~~, and answering any questions.

Possible risks and discomforts No extraordinary risks are anticipated in the present study. The risk is expected to be no more than that of every day's life. Your name and any other personal information will not be saved. ~~connected to the data collected from you.~~

Possible benefits You will be given $15 CAD for compensation for the study even if you spend more or less than the expected time in it. You will be paid the full compensation even if you do not complete the study and plan to withdraw. In addition, your participation will be greatly appreciated, and we expect that it will help us to evaluate accuracy of Machine

Learning system which can classify documents automatically into topics and to learn better how users can interact with user interface of the visualization system to do the set of tasks given as part of visualization system testing. You can also download your classified topics in a digital document after the study is done.

What you will be asked to The users will ~~sign the~~ provide the consent using online consent form~~.~~, This is expected to take 10 minutes. ~~complete the screening questionnaire, demographic questionnaire followed by tutorial on using the software and user interface.~~ Further, introduction and training will be provided to be familiar with the system and the user interface ~~These processes are~~ This step is also expected to take ~~15~~ 10 minutes. In the end, the participants fill out the post-condition questionnaire. It is expected that the interaction with the system will take about 30 minutes. The user study is implemented as a web page and whole processing, from uploading papers to downloading the classified topics and interacting with the visualization system, are online.

- **You** ~~will sign~~ provide consent by our online ~~the~~ **consent form**~~.~~
- ~~You will complete a screening questionnaire.~~
- ~~You will complete a demographic questionnaire.~~
- ~~You will be given a tutorial on how to use the software.~~
- You will be given a practice session/tutorial to use the ~~software~~ system and user interface.
- You will be given user Ids, set of questions that will make you interact with our system followed by the evaluation (post-condition) questionnaire. This complete step is online.
- You will perform the interaction with the system in a quiet room in the Computer Science Faculty. However, you have to bring your own laptop for this study.
- You will submit the post-study questionnaire and get your compensation. ~~If you wish to participate in the study then please:~~
- ~~Sign this consent form, and Give it to the principal investigator.~~

~~The principal investigator will keep the original copy of this consent form.~~

Confidentiality Your name and address will not be required when answering any of the questions. We store a log of participant's time for interacting with the Excel sheet and time for interacting with the user interface of the Visualization system along with DOI (Digital Object Identifier) of the research paper that you bring with you, the classified topics extracted by the system, the edit to the topics by the "expert" (participant) who bring the paper in and the number of correctly classified topics as evaluated by the participant. No participant can see or access documents from other participants. All of your data will be treated confidentially and stored in a secure location. All data gathered from this study may potentially be used in

publications. The data will be retained for five years after publication and then destroyed. Your personal documents will not be stored in the system. Classified topics of an academic research paper will be available for you to download after study. Since the study population is less (10 participant), identification to the level of the research group/interest might be possible, but it is not easy to identify an individual due to the coarse-grained nature of the topics (40 topics for all of Computer Science). The emails that participants send to the PI for showing interest to take part in the study will also be deleted after the study. We are not storing names or any personal information about participants. We are only storing user IDs. All data that we collect as part of the study will not be accessible by any participant.

Conflict of Interest No identifying information is being stored and there is no conflict of interest between anyone.

Possible Follow-ups There's no follow up required for any further study.

Questions Please feel free to ask the principal investigator (Deepak Munjal) about anything to do with the study before (or after) you give consent to participate.

Signature ~~By signing below~~ By choosing the "Yes" option below, you are agreeing to the following statements:

· I understand that I may discontinue my participation at any point in the study.

· I have read and understood the procedure and nature of this study.

· I have had a chance to ask any questions I have about the study, and they have been answered.

· I am aware that all research material will be kept confidential.

· My quotations might be used in a publication with anonymity.

· I understand that if I have any complaints about the experiment that I may contact:

> Catherine Connors,
> Human Research Ethics/Integrity coordinator,
> Office of Research Ethics, Dalhousie University.
> Email: {catherine.connors@dal.ca}
> Phone: (902) 494-1462

I hereby consent to voluntarily take part in this study.

O Yes
O No

Name:_____

Signature:_____

Date:_____

Additional optional permissions

Please indicate your selections and initial or sign each item below.

Study Results

_____ I want / do not want to receive the study results.

Email:_____

I grant permission for Ddirect publication of quotations from my post-study questionnaire _____ I grant / do not grant permission forto Deepak Munjal. to publish quotations from my post-study questionnaire.

O Yes
O No

Appendix C – Screening Questionnaire

Identification number: _____

1. At what level do you think your understanding of written English is?
   - Excellent
   - Very good
   - Good
   - Acceptable
   - Bad
   - Very bad
   - None

2. What is the highest level of education you have completed?
   - Little or no formal education
   - High school or equivalent
   - College or university
   - Master
   - Doctoral
   - Post-Doctoral

3. How often do you read newspapers or magazines?
   - Every day
   - Once two days
   - Once four days
   - Once a week
   - Once a month
   - Once a year
   - Never

4. How often do you read academic papers?
   - Every day
   - Once two days
   - Once four days
   - Once a week
   - Once a month
   - Once a year
   - Never

5. What is your primary area of study?

- Computer Science
- Information Technology
- Internetworking
- Other_____

Appendix D–Demographic Questionnaire

Identification number: _____

Gender: _____Male

_____Female

_____Other

1. How long have you used a computer?
   - Less than 1 day
   - 1 day to less than 1 week
   - 1 week to less than 1 month
   - 1 month to less than 6 months
   - 6 months to less than 1 year
   - More than 1 year

2. On the average, how much time do you spend per week on a computer?
   - Less than one hour
   - One to less than 4 hours
   - 4 to less than 10 hours
   - 10 to less than 20 hours
   - 20 to less than 40 hours
   - Over 40 hours

3. How often do you use interactive user interfaces such as searching queries in search boxes, clicking buttons etc?
   - Extremely often
   - Very often
   - Often
   - Not often
   - Seldom
   - Never

4. How comfortable are you at using interactive user interface?
   - Extremely comfortable
   - Very comfortable
   - Comfortable
   - Uncomfortable
   - Very uncomfortable

Extremely uncomfortable

5. If your major is computer science or information technology, how familiar with the following area:

6. Software, software engineering

- · Very well
- · Well
- · Neutral
- · Not well
- · Not well at all
- · Not Applicable

Programming languages

- · Very well
- · Well
- · Neutral
- · Not well
- · Not well at all
- · Not Applicable

Information systems, e.g., database and data mining

- · Very well
- · Well
- · Neutral
- · Not well
- · Not well at all
- · Not Applicable

Computing methodologies, e.g., artificial intelligence

- · Very well
- · Well
- · Neutral
- · Not well
- · Not well at all
- · Not Applicable

Appendix E D – Visualization Component Test Quizzes

Identification number: _____

Answers to these questions are not completely correct or incorrect. You can answer what you think is the most relevant answer:

S. No	Question	Anwer
1	You are staying in ~~Halifax~~New York and looking for a Speaker who can deliver a lecture on ~~machine learning~~Artificial Language/Machine Learning in ~~Halifax~~New York itself. Considering the limited budget, whom would you prefer?	
2	How many Computer Graphics speakers need to be recruited in Germany?	
3	Which topic is trending more from July 2017 to March 2018 in Canada?	
4	Which country has the maximum number of seminars on Human-Computer Interaction in the year 2017?	
5	How many seminars occurred on Text mining in New York in the year 2017?	

Appendix EF – Post-condition Questionnaire

Identification number: _____

Please respond to the following statements using the given scale (circle response):

		1	2	3	4	5
1.	You have sufficient background about the document classification.	*Strongly Disagree*	*Somewhat Disagree*	*Neutral*	*Somewhat Agree*	*Strongly Agree*
2.	It is easy to identify all the topic(s)/categories of a research document manually by reading it?	*Strongly Disagree*	*Somewhat Disagree*	*Neutral*	*Somewhat Agree*	*Strongly Agree*
3.	The system has made it easy to classify a large set of research documents into topics.	*Strongly Disagree*	*Somewhat Disagree*	*Neutral*	*Somewhat Agree*	*Strongly Agree*
4.	The set of topics predicted by the system is far from what you desired?	*Strongly Disagree*	*Somewhat Disagree*	*Neutral*	*Somewhat Agree*	*Strongly Agree*
5.	The user interface of the machine learning system is intuitive and easy to use.	*Strongly Disagree*	*Somewhat Disagree*	*Neutral*	*Somewhat Agree*	*Strongly Agree*
6.	The machine learning system is fast enough while classifying into topics.	*Strongly Disagree*	*Somewhat Disagree*	*Neutral*	*Somewhat Agree*	*Strongly Agree*
7.	The visualization system has made it simple to interact with the database.	*Strongly Disagree*	*Somewhat Disagree*	*Neutral*	*Somewhat Agree*	*Strongly Agree*
8.	The user interface of the visualization system is intuitive and easy to use.	*Strongly Disagree*	*Somewhat Disagree*	*Neutral*	*Somewhat Agree*	*Strongly Agree*
9.	The visualization system is fast enough while interacting with the database.	*Strongly Disagree*	*Somewhat Disagree*	*Neutral*	*Somewhat Agree*	*Strongly Agree*

		1	2	3	4	5
10.	The number of options provided in The user interface for the visualization system is enough.	*Strongly Disagree*	*Somewhat Disagree*	*Neutral*	*Somewhat Agree*	*Strongly Agree*
11.	The overall system requires high technical skills.	1	2	3	4	5
		Strongly Disagree	*Somewhat Disagree*	*Neutral*	*Somewhat Agree*	*Strongly Agree*
12.	With proper documentation, you want to use the software in future, if required?	1	2	3	4	5
		Strongly Disagree	*Somewhat Disagree*	*Neutral*	*Somewhat Agree*	*Strongly Agree*
13.	How much time (ai n hours) did you spent in interacting with the system?	*Less than 1*	*1-2*	*2-3*	*3-4*	*More than 5*
14.	How likely would you recommend this tool to anyone else, looking to classify a document into topics?	1	2	3	4	5
		Strongly Disagree	*Somewhat Disagree*	*Neutral*	*Somewhat Agree*	*Strongly Agree*

15. Please give us more comments about the system:

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

16. Is there any functionality that you expect to be included but was not available?

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Appendix FG - Participant Payment Receipt

My signature below confirms that I received a sum of $15 (CDN) cash from Deepak Munjal as an honorarium payment for participating in the "Visual Analytics of Research Community Expertise in Space and Time" research thesis.

I understand this honorarium is taxable income and it is my responsibility to declare it on my income tax return as Dalhousie University will not be issuing a T4A for this payment.

Name (please print): _____

Signature: _____

Date: _____