Clustered Layout Word Cloud for User Generated Online Reviews

Ji Wang

Thesis submitted to the Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Computer Science and Applications

Christopher L. North, Chair

Yong Cao

Naren Ramakrishnan

October 25, 2012

Blacksburg, Virginia

Keywords: Word Cloud, Text Visualization, User Study

VirginiaTech

MEMORANDUM

DATE: April 5, 2012

TO: Christopher L. North, Ji Wang

FROM: Virginia Tech Institutional Review Board (FWA00000572, expires May 31, 2014)

PROTOCOL TITLE: Evaluation of Context Preserving Tag Cloud for User Generated Online Reviews

IRB NUMBER: 12-348

Effective April 4, 2012, the Virginia Tech IRB Chair, Dr. David M. Moore, approved the new protocol for the above-mentioned research protocol.

This approval provides permission to begin the human subject activities outlined in the IRB-approved protocol and supporting documents.

Plans to deviate from the approved protocol and/or supporting documents must be submitted to the IRB as an amendment request and approved by the IRB prior to the implementation of any changes, regardless of how minor, except where necessary to eliminate apparent immediate hazards to the subjects. Report promptly to the IRB any injuries or other unanticipated or adverse events involving risks or harms to human research subjects or others.

All investigators (listed above) are required to comply with the researcher requirements outlined at http://www.irb.vt.edu/pages/responsibilities.htm (please review before the commencement of your research).

PROTOCOL INFORMATION:
Approved as: **Expedited, under 45 CFR 46.110 category(ies) 7**
Protocol Approval Date: **4/4/2012**
Protocol Expiration Date: **4/3/2013**
Continuing Review Due Date*: **3/20/2013**
*Date a Continuing Review application is due to the IRB office if human subject activities covered under this protocol, including data analysis, are to continue beyond the Protocol Expiration Date.

FEDERALLY FUNDED RESEARCH REQUIREMENTS:
Per federally regulations, 45 CFR 46.103(f), the IRB is required to compare all federally funded grant proposals / work statements to the IRB protocol(s) which cover the human research activities included in the proposal / work statement before funds are released. Note that this requirement does not apply to Exempt and Interim IRB protocols, or grants for which VT is not the primary awardee.

The table on the following page indicates whether grant proposals are related to this IRB protocol, and which of the listed proposals, if any, have been compared to this IRB protocol, if required.

Invent the Future

Date*	OSP Number	Sponsor	Grant Comparison Conducted?

*Date this proposal number was compared, assessed as not requiring comparison, or comparison information was revised.

If this IRB protocol is to cover any other grant proposals, please contact the IRB office (irbadmin@vt.edu) immediately.

cc: File

Informed Consent for Participant of Investigative Project

Virginia Polytechnic Institute and State University

Title of Project: Evaluation of Context Preserving Tag Cloud for User Generated Online Reviews

Principal Investigator: Dr. Chris North
Co-Investigators: Ji Wang

I.   THE PURPOSE OF THIS RESEARCH/PROJECT

You are invited to participate in a study of text visualization. This study researches the ways people use our new context preserving tag cloud technique to analyze the online reviews data from Yelp.com. This study involves experimentation for the purpose of evaluating and improving the context preserving tag cloud visual analytics system.

II.   PROCEDURES

You will be asked to analyze the online reviews from Yelp.com for restaurants by using three different approaches we provide: normal plain text, normal tag cloud and our context preserving tag cloud. Your role in these tests is to be an evaluator of different approaches. We are not evaluating you or your performance in any way; you are helping us to evaluate our system. All information that you help us attain will remain anonymous. The amount of the information you get and the time you take to do each task will be measured. You may be asked questions during and after the evaluation in order to clarify our understanding of your evaluation.

You will also be asked to fill out questionnaires relating to your background with the experience of online reviews reading and to collect your feedback.

The session will last around one hour. The analysis tasks are not very difficult, but you are welcome to take rest breaks as needed. You may also terminate your participation at any time, for any reason.

You will be given full instructions and a chance to practice each approach we provide, prior to using it for an analysis task. If anything is unclear, be sure to ask us questions.

III.   RISKS

In this study, there are no known risks. In case of exhaustion you are encouraged to take a break. The task does not require long time continuous activity.

IV.   BENEFITS OF THIS PROJECT

Your participation in this project will provide information that may be used to improve our context preserving text visualization. You may receive a synopsis summarizing this research when completed. Please leave a self-addressed envelope with the experimenter and a copy of the results will be sent to you.

You are requested to refrain from discussing the evaluation with other people who might be in the candidate pool from which other participants might be drawn.

V.   EXTENT OF ANONYMITY AND CONFIDENTIALITY

The results of this study will be kept strictly confidential. Your written consent is required for the researchers to release any data identified with you as an individual to anyone other than personnel working on the project. The

information you provide will have your name removed and only a subject number will identify you during analyses and any written reports of the research. So, confidentiality and anonymity are guaranteed.

VI.   COMPENSATION

You will be paid US$ 30 for the study.

VII.   FREEDOM TO WITHDRAW

You are free to withdraw from a study <u>at any time</u> without penalty. There may be circumstances under which the investigator may determine that you should not continue as a subject. You will be thanked for you time.

VIII.   APPROVAL OF RESEARCH

This research has been approved, as required, by the Institutional Review Board for projects involving human subjects at Virginia Polytechnic Institute and State University, and by the Department of Computer Science.

IX.   SUBJECT'S RESPONSIBILITIES AND PERMISSION

I voluntarily agree to participate in this study, and I know of no reason I cannot participate. I have read and understand the informed consent and conditions of this project. I have had all my questions answered. I hereby acknowledge the above and give my voluntary consent for participation in this project. If I participate, I may withdraw at any time without penalty. I agree to abide by the rules of this project

_____   _____
Signature   Date

_____   _____
Name (please print)   Contact: phone or address or

                                                                                          _____
                                                                                          email address (OPTIONAL)

Should I have any pertinent questions about this research or its conduct, and research subjects' right, and whom to contact in the event of a research-related injury to the subject, I may contact:

Investigators: Dr. Chris North (Faculty Advisor),   Ji Wang
                              north@cs.vt.edu   wji@cs.vt.edu
                              Department of Computer Science
                              2202 Kraft Drive
                              Blacksburg, VA. 24060
                              Phone: 540-231-7409

Review Board: David M. Moore, Office of Research Compliance, CVM Phase II (0442) 231-4991

cc: the participant, Dr. North

Title: Invitation to participate in visual analytics for online reviews study

To: HCI-STUDENTS, cs grad students listserv, cs undergrad students listserv

We are looking for participants to take part in our study "Evaluation of Context Preserving Tag Cloud for User Generated Online Reviews". The study will include: (1) use our Context Preserving Tag Cloud and normal Tag Cloud to analyze online reviews; (2) answer a few questionnaires.

The study will take approximately ONE hour and will be located at 1156 Knowledge Work II. If you are interested in participating, please reply to Ji Wang (wji@cs.vt.edu) to schedule a time.

Every participant will get $ 30 for the compensation.

Please note that you must be at least 18 years old and be able to read and analyze online reviews on Windows system for about one hour to participate in this study.

This study has been approved by the Institutional Review Board of Virginia Tech.

Thank you,

Ji Wang

User Pre-Study Questionnaire

Please help us to categorize our user population by completing the following items.

Gender (circle one):   Male   Female

Age: _____

Are you an English native speaker?

Yes   No   If No, please provide your first language: _____

Occupation (if student, indicate graduate or undergraduate):

_____

Major / Area of specialization (if student): _____

How often do you use online reviews resource to support your decision making in dinning, shopping, entertainment (like movie and video games), travel (like hotel and flight) and etc. (circle only one)?

   a. Not at all
   b. Once a month
   c. Once a week
   d. Several times a week
   e. Daily

If you use online review resource, which website do you use regularly (circle all that apply)?

   a. Amazon.com
   b. IMDB
   c. Yelp.com
   d. Trip Advisor
   e. Google Places
   f. Other(s) _____

Which types of activities do you use online reviews usually (circle all that apply)?

   a. Dining
   b. Shopping
   c. Entertainment

d. Travel
e. Other(s) _____

Have you used the online reviews from Yelp.com to find a dinning place before (circle one)?

Yes   No

       If **yes**, how often do you use the online reviews from Yelp.com?

          a. Several times a year
          b. Once a month
          c. Once a week
          d. Several times a week
          e. Daily

Have you written any online reviews before (circle one)?

Yes   No

       If **yes**, how often do you write online reviews?
          a. Several times a year
          b. Once a month
          c. Once a week
          d. Several times a week
          e. Daily

       If **yes**, can you provide what websites/platforms did you write?

       _____

Do you have any suggestions about the design of current online reviews website or platform?

_____

_____

_____

Technique Questionnaire I

User ID: _____   Analysis Technique: _____

Test 1: Decision Making Test
Which one will you choose or would like to suggest to your friends?

A   **B**

Please explain the criteria you used to determine the above choice.

```



```

Which one will you choose or would like to suggest to your friends?

A   **B**

Please explain the criteria you used to determine the above choice.

```



```

Test 2: Decision Making Test
Which one will you choose or would like to suggest to your friends?

A   B

Please explain the criteria you used to determine the above choice.

[   ]

Which one will you choose or would like to suggest to your friends?

A   B

Please explain the criteria you used to determine the above choice.

[   ]

Test 3: Decision Making Test
Which one will you choose or would like to suggest to your friends?

A   **B**

Please explain the criteria you used to determine the above choice.

Which one will you choose or would like to suggest to your friends?

A   **B**

Please explain the criteria you used to determine the above choice.

Please rank the analysis techniques in order of preference where 1 is most-preferred, and 3 is least preferred, and each number (1 to 3) is used only once.

___ Normal Plain Reviews
___ Normal Tag Cloud
___ Context Preserving Tag Cloud

Please explain the criteria you used to determine the above order.

Technique Questionnaire II

User ID: _____   Analysis Technique:

_____

Test 4: Feature Finding Test
Food Quality:
Please provide a summary and explain the features you have found

Service Quality:
Please provide a summary and explain the features you have found

Technique Questionnaire 2.2

User ID: _____   Analysis Technique:
_____

Test 5: Feature Finding Test
Food Quality:
Please provide a summary and explain the features you have found

Service Quality:
Please provide a summary and explain the features you have found

Technique Questionnaire 2.3

User ID: _____   Analysis Technique:
_____

Test 6: Feature Finding Test
Food Quality
Please provide a summary and explain the features you have found

Service Quality
Please provide a summary and explain the features you have found

Please rank the analysis techniques in order of preference where 1 is most-preferred, and 3 is least preferred, and each number (1 to 3) is used only once.

___ Normal Plain Reviews
___ Normal Tag Cloud
___ Context Preserving Tag Cloud

Please explain the criteria you used to determine the above order.

For Context Preserving Tag Cloud

This technique was easy to learn.

 Strongly Disagree   Disagree   Neutral   Agree   Strongly Agree

The technique felt natural/intuitive.

 Strongly Disagree   Disagree   Neutral   Agree   Strongly Agree

This technique was appropriate for the given task.

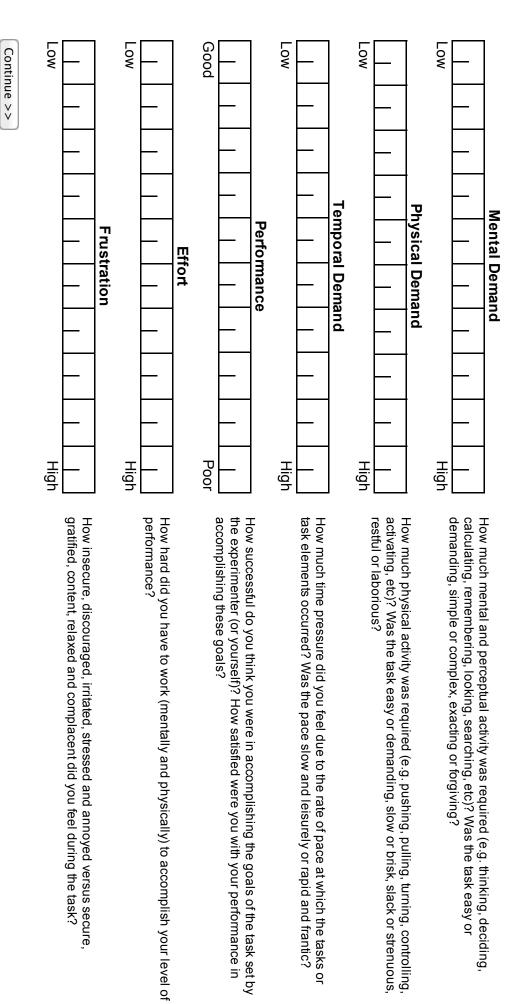 Strongly Disagree   Disagree   Neutral   Agree   Strongly Agree

Analysis in this way was fun.

 Strongly Disagree   Disagree   Neutral   Agree   Strongly Agree

If you had any challenges or problems using this technique, please describe:

Task Questionnaire - Part 1

Click on each scale at the point that best indicates your experience of the task

Mental Demand

Low | | | | | | | | | | | | | | | | | | | High

How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc)? Was the task easy or demanding, simple or complex, exacting or forgiving?

Physical Demand

Low | | | | | | | | | | | | | | | | | | | High

How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Temporal Demand

Low | | | | | | | | | | | | | | | | | | | High

How much time pressure did you feel due to the rate of pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

Performance

Good | | | | | | | | | | | | | | | | | | | Poor

How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

Effort

Low | | | | | | | | | | | | | | | | | | | High

How hard did you have to work (mentally and physically) to accomplish your level of performance?

Frustration

Low | | | | | | | | | | | | | | | | | | | High

How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Continue >>

User Post-Study Questionnaire

Please rank the analysis techniques in order of preference where 1 is most-preferred, and 3 is least preferred, and each number (1 to 3) is used only once.

___ Normal Plain Text Reviews
___ Normal Tag Cloud
___ Context Preserving Tag Cloud

Please explain the criteria you used to determine the above order.

How likely would you be to choose Context Preserving Tag Cloud technique over normal plain text online review in your daily life (circle one)?

| Very Likely | Somewhat Likely | Neutral | Somewhat Unlikely | Very Unlikely |

Why?

Please use the following space to leave any comments or suggestions about our study.

Thank you for participating in the study!