Helping students find answers: Algorithmic interpretation of student questions

Jeremy Herbert[a], Elliot Smith[a], Carl Reidsema[a], Lydia Kavanagh[a],
University of Queensland[a]
Corresponding Author's Email: jeremy.herbert@uqconnect.edu.au

Structured Abstract

CONTEXT
In discussion boards with a large number of participants, students often struggle with information overload and are more likely to post short answers and minimise involvement (Jones, Q., Ravid, G., & Rafaeli, S., 2004). Additionally, questions asked by students are often duplicated, making answers difficult to find as they dilute the available information on a particular topic. As more and more university courses transition to a flipped classroom model with diverse cohorts that do not have ready access to face-to-face assistance, this problem will be exacerbated. It is important that systems are put into place that can help students manage the large amount of information that can be shared and stored in discussion groups.

PURPOSE OR GOAL
Design and develop a system where students can post questions that can be answered by both teaching staff and students. The system will be used to:

- Ensure students equal access to teaching staff
- Allow students to manage their own learning
- Increase the efficiency of teaching staff

APPROACH
A new discussion system (Casper) has been developed with an integrated algorithm for comparing the similarity of question based on their use of distinct words (words which are used rarely over the whole text database) and other metrics such as number of views. When a student attempts to ask a question, the question title is used to perform a query and questions that contain similar distinct words are returned. The user is then displayed with an indication that their question may have already been asked and a link to redirect the student to the question. The system will be deployed in two compulsory, team- and project-based courses at XXXX for one semester and any use of the suggested links will be recorded. Students will be surveyed in a focus group format during the semester to collect feedback about the system.

ACTUAL OR ANTICIPATED OUTCOMES
It is expected that the system will improve student feedback giving greater access to teaching staff to assist them in managing their own learning.

CONCLUSIONS/RECOMMENDATIONS/SUMMARY
The likely conclusion of this study will be that the suggestions are utilized by students to reduce information overload. If this is the case, it will be important to disseminate this tool across other courses.

REFERENCES (OPTIONAL)
Jones, Q., Ravid, G., & Rafaeli, S. (2004). Information overload and the message dynamics of online interaction spaces: A theoretical model and empirical exploration. *Information systems research.*

KEYWORDS
eLearning, discussion, search

Introduction

Online discussion is considered a necessity in modern university courses because it provides students with an avenue to share and debate ideas with their colleagues. However, it can become unwieldy and less useful with a large numbers of participants as relevant information becomes more difficult to find amongst the large volumes of discussion threads. When this occurs, students are likely to suffer from 'information overload', or "a state of having too much information to digest, with resulting negative effects on performance or wellbeing" (Strother, Ulijn, & Fazal, 2012). In a study of hundreds of online newsgroups, researchers found that as the number of users increased, individuals began to respond to only shorter messages, write shorter responses and reduce their overall participation in the group (Jones, Ravid, & Rafaeli, 2004). This represents a problem for current university online discussion tools as they have been modelled after this threaded discussion system. In an attempt to counteract the effects of information overload on students, a new online discussion tool named Casper was developed with a goal of collecting metadata about discussions, and using this information to improve student feedback through minimizing information overload.

Many online services need to build in features that help users manage information overload once their user-base grows otherwise they risk reaching a 'virtual population limit' which is defined as the maximum number of users that can effectively access and contribute information to the system. A common and effective method to do this is to computationally process the information being accessed/created and provide the user with a suggestion of actions or data that may aid them in achieving their perceived task. A well-known example of this is friend suggestion on Facebook. When the user enters a search term to try and find a new friend, Facebook does not simply perform a text query with the entered name. Instead, it traverses the social graph surrounding the user and tries to find people with a name similar to the query that are closely connected to the user; after all, one is more likely to meet a friend-of-a-friend in day to day life than an unknown person on the other side of the world (Roth et al., 2010). In this way, Facebook reduces the cognitive load associated with identifying people in a large number of listings, and hence makes the body of information appear to the user as much smaller than it actually is. This automated filtering of information to reduce the size of a dataset is the key technique that forms the basis for online search engines, online shopping recommendation tools and many more online services (Berghel, 1997).

The examples presented thus far all require active input from the user; however, information suggestions do not need to be the result of an intentional action. Facebook will automatically suggest new friends to a user on the side of the page by monitoring which pages are accessed with respect to their own social sub graph. Google's Gmail will tailor advertisements to relate to the content of the user's emails. These passive suggestions are useful as they can be used to encourage a user to behave in a way that is beneficial to the service.

The aim of this study was to trial the perceived usefulness of information suggestion features in the context of a university course discussion tool. A passive suggestion feature was implemented in the discussion tool that processed new questions as they were being asked and provided an indication to the user if another student had already asked a similar question. After deploying this tool in undergraduate courses at The University of Queensland, students were surveyed in order to determine their opinions on this specific feature.

Method

The Casper question and answer tool was developed at the University of Queensland in 2013. It was modelled after the popular online discussion system 'StackExchange' and was designed to seamlessly integrate with the legacy Blackboard Learning Management System (LMS) as a drop-in replacement for the built-in threaded discussion tool. Blackboard has

been used at the university for many years and represents the core framework for dissemination of online learning materials.

Casper was deployed in two undergraduate engineering courses for semester 2, 2013.The enrolment count for these courses was approximately 1200 students and 300 students respectively. Both courses involve a significant amount of teamwork activities and focus on student-driven learning.

Upon visiting Casper, students are presented with the list of questions that have currently been posted, a list of popular/recent tags and a navigation bar that gives them the opportunity to ask a question or search for a question/answer by text.

They are then presented with a form that allows them to ask a question. After they have entered a title, the system performs a text search with the title as the query and displays any similar questions to the user (Figure 1). The text search algorithm was identical to the one used for other search functionality in the tool.

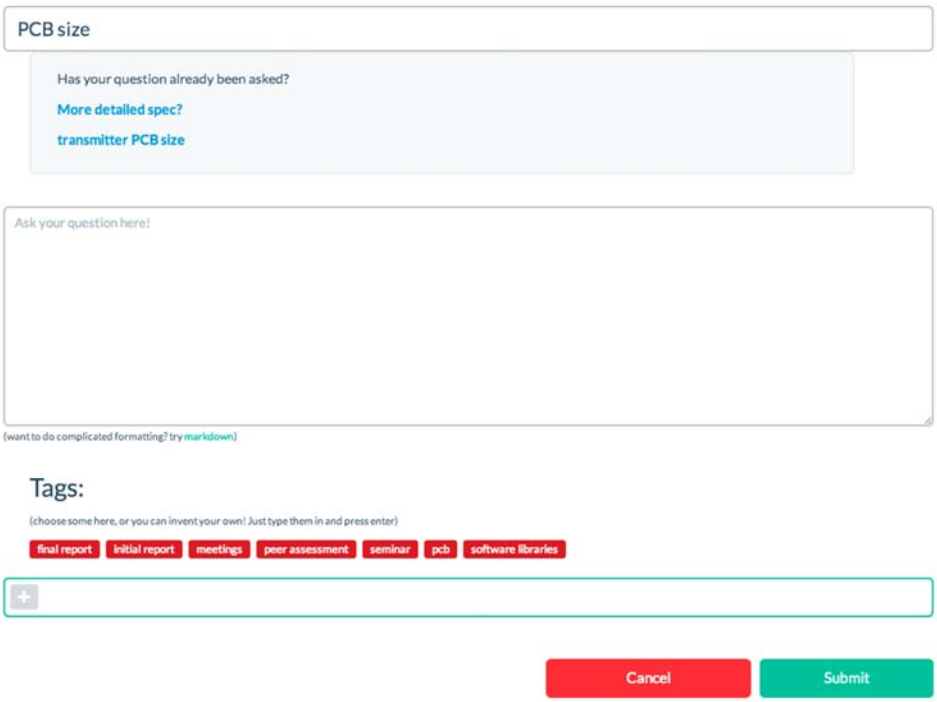

Figure 1: Question form with suggestions visible below the title field

To determine the perceived usefulness of these suggestions by students, an online survey was sent via Blackboard to all enrolled students. Students were only able to respond to questions on a five point Likert scale (except for one yes/no question and one freeform comment). Three questions related to the question suggestion feature were in the survey:

1. Has Casper suggested similar questions while you are asking your own? (response was yes or no only).
2. If yes [in reference to the previous question], Were the suggestions were helpful?
3. Suggesting similar questions is a useful feature for the tool to have.

The first of these questions was used to remove any erroneous responses if the students had not actually used the suggestion feature. However, all responses were counted for question three whether the student had used the suggestions or not.

The survey was sent out after one month of the semester, and responses were collected over a five-day period.

Results

A total number of 171 students responded to the survey over the five-day period. The responses provided are outlined in the figures below.

Statement: If yes [referring to the previous question on whether the tool had provided suggestions or not], the suggestions were helpful:

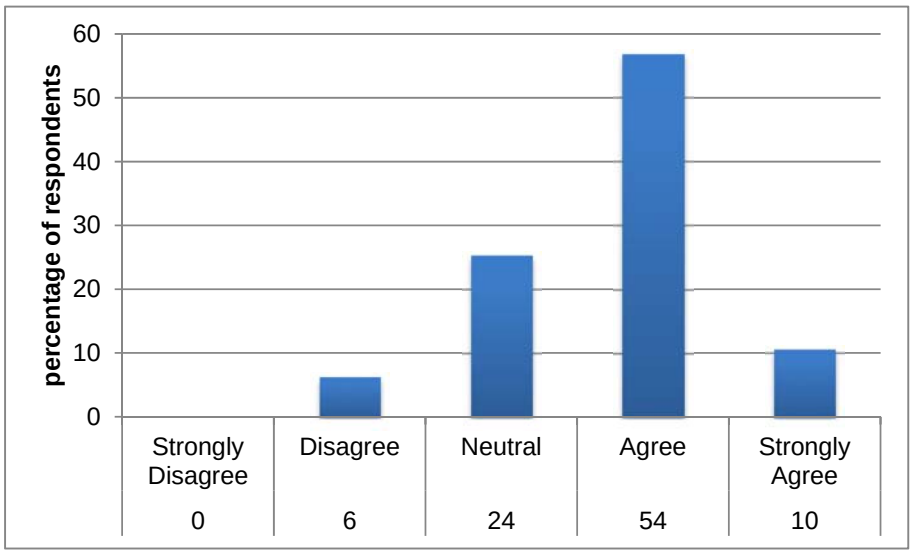

Figure 2: Responses to question two in the survey. The number at the bottom of each column is the total number of responses in the category.

Statement: Suggesting similar questions is a useful feature for the tool to have:

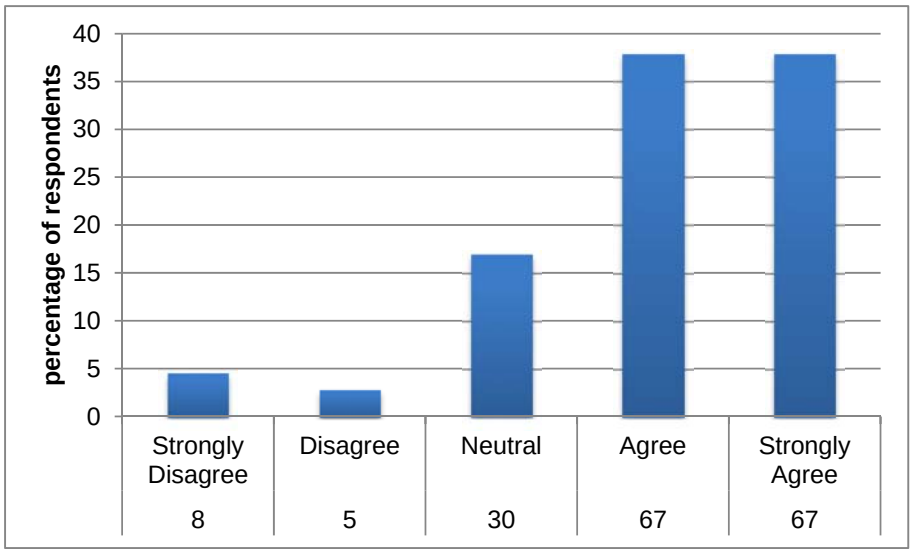

Figure 3: Responses to question three in the survey. The number at the bottom of each column is the total number of responses in the category.

Discussion

Student perceptions of the question suggestions were strongly positive. From the information in Figure 1, 67% of students felt that the suggestions that the tool had provided them with were helpful. This number suggests that after only a month of the semester, students were already ineffective at finding the information that they needed in the discussion boards. However, the majority of this group did not 'strongly agree', but rather only agreed that the question suggestions were helpful. This may indicate that further tuning of the suggestion algorithm is needed to ensure that only highly relevant questions are suggested.

Students were also overwhelmingly in agreement (75% of respondents) that the suggestion feature was a useful feature for the tool to have. This may be because they had either already been helped by the feature, or perhaps they had previously struggled with the effects of information overload in other online discussion tools.

A small number of students indicated that the tool was not helpful and that they would not like the tool to perform an automated question suggestion. To account for these students, a discussion tool should allow students to turn off any information suggestion features. However, given the strong support for the features, they should be enabled by default.

It is also notable that students found the suggestion feature useful despite the fact that the exact same algorithm was used if keywords were entered in the search box that was visible on every page. This could be a result of a lack of awareness of how to search correctly, or perhaps because students did not feel that searching would be useful for finding answers to their question. Either of these represents a problem, as this indicates that students were either ill prepared to perform, or otherwise unaware of the benefits of, proper information searching and retrieval techniques.

Conclusion

Online discussion systems are now a key instrument in the toolbelt of a university educator. They provide an easy and effective way for students to collaborate in courses as they provide an avenue for students to have contact with a greater fraction of their cohort than if they were to speak to their colleagues individually. However, we must consider that adapting our teaching methodologies to much larger class sizes without adapting our traditional educational tools is likely to present us with exacerbated problems like information overload.

Tools like Casper are the first step toward solving this problem, in that they record a significant amount of data for research purposes, but also because they integrate features that reduce the perceived amount of data available to a student and thus aid them in dealing with information overload. Many of these techniques are commonplace in online services, and it remains somewhat of a mystery to the authors as to why they have not already been integrated into the standard LMS packages; it is clear from the survey data that students value these features in online discussion tools.

There is already ample evidence to suggest that information overload is a significant problem for students in these environments, and the results of this study further indicate that this is a problem that needs more focus from educators and the developers of educational software. With further research, it is hoped that we can build tools that aid students in reducing the amount of irrelevant information they need to process, thus relieving some of the burden of information overload. At that point, students can more actively engage their original and more appropriate problem: undertaking discussion with their colleagues to develop a deeper and more effective understanding of the course material.

References

Berghel, H. (1997). Cyberspace 2000: dealing with information overload. *Communications of the ACM*, *40*(2), 19–24. doi:10.1145/253671.253680

Jones, Q., Ravid, G., & Rafaeli, S. (2004). Information overload and the message dynamics of online

interaction spaces: A theoretical model and empirical exploration. *Information systems research*.

Roth, M., Ben-David, A., Deutscher, D., Flysher, G., Horn, I., Leichtberg, A., et al. (2010). Suggesting friends using the implicit social graph (p. 233). Presented at the the 16th ACM SIGKDD international conference, New York, New York, USA: ACM Press. doi:10.1145/1835804.1835836

Strother, J. B., Ulijn, J. M., & Fazal, Z. (2012). *Information Overload*. John Wiley & Sons.

Copyright statement