# Answering Student Programming Questions using Domain-specific Search

Kelvin Meyer and Hussein Suleman[0000−0002−4196−1444]

Department of Computer Science
University of Cape Town
Private Bag X3, Rondebosch
South Africa
MYRKEL002@myuct.ac.za, hussein@cs.uct.ac.za

**Abstract.** Discussion forums are commonly used in online learning environments for teaching programming, to create a platform for students to discuss course content. This platform of interaction is not without its challenges, as students regularly repeat questions that others have asked, both within and across offerings of a particular course. If past answers can be reliably provided to students, it eliminates the need for repetition and provides students with immediate assistance. This study investigates an approach to enable this through the addition of a search feature that indexes and queries discussion forum messages from a previous year to answer student questions. In particular, the paper presents a comparison of different ranking approaches based on the exploitation of domain-specific features of a social discussion forum on a learning management system, in particular, the authority of respondents. Results show that information retrieval can yield relevant answers to students in a programming course within the first 3-5 results, with some improvement in the outcomes when the social notion of authority in exploited.

**Keywords:** discussion forum, information retrieval, search, social discussion

## 1 Introduction

Discussion forums are a common tool for students and teachers to use in online learning environments such as learning management systems. They offer a new platform for communication that improves the learning process for the students who participate in the forum [5] and adds a new dimension to the course in conjunction with traditional teaching methods of lectures and tutorials. The first year computer science courses at University X make use of a basic threaded social discussion forum that students in the course can use to post questions they have about the course content or any other matter relating to the course as a whole. Social discussion forums go beyond discussion forums by incorporating elements commonly found in social networks; in the system that is the focus of this study, there are identities and groups, notifications and likes [15].

The threaded nature of the forum allows lecturers, tutors and other students to answer questions and then have focused discussions [12] while other questions can be asked and then responded to in their own threads. This forum is used by the students for topics ranging from questions about course content to general course administration questions to other common questions many first year students have. The discussion forum functions well as a platform for communication amongst students as it is always available and the asynchronous nature of such a discussion forum also allows lecturers and tutors to provide feedback and other forms of assistance to students when they are able to do so. While the students in the courses change every year, the lecturing staff change less frequently. This leads to questions from students in the discussion forum being repeated from year to year and sometimes even within the same year. This can be frustrating for the lecturing staff who have to repeatedly answer the same questions. More critically, if the answer is already available in a previous offering of the course, there is need to answer the same question again. Further, this answer can be provided to students immediately.

This paper argues that a search feature would allow students to look for previous answers before posting new questions, but also that a domain-specific search system, using variable boosted queries by authority of posters, would make better use of the nature of social discussion data to provide more relevant answers to users.

## 2 Information Retrieval

Information Retrieval [8] is the area of Computer Science concerned with the development of search engines and related technologies. A search engine is fundamentally comprised of an indexing system that processes a collection of data and creates an index; and a querying system that submits queries and returns a list of ranked results to the user.

Ranked retrieval [9] is what makes such systems fundamentally different from traditional database systems. In a traditional database, an item of data either matches or does not match a query. In a traditional information retrieval system, every item of data matches a query to a certain degree, and this degree can be used to sort results in order of probable relevance. Common formulations for this matching function include a combination of the number of occurrences of a term in a document and the rarity of the term.

Domain-specific search goes beyond general information retrieval algorithms to incorporate and exploit aspects of the data specific to an application areas. For example, patent search [11] will incorporate structural information into the matching function so some aspects of the patents are deemed to be more important than others.

Information retrieval algorithms are evaluated by comparing algorithms using a set of well-established metrics calculated on whether or not a user deems a result to be relevant to a query (for every query and every result) [10] [8]:

- Recall, which indicates how many of the relevant results have been found

– Precision, which indicates how many of the results that have been found are relevant
– Mean Average Precision, which takes the average precision of the set {first 1 result, first 2 results, ... first n results}
– Normalized Discounted Cumulative Gain, which takes a weighted average that favours results at the top of the list

## 3 Related work

The use of information retrieval techniques for many Web search tasks is well understood [13], however, online discussion forums represent a different challenge in that they tend to follow conversational style of group interactions that differs significantly from the Web 1.0 [14] approach of a single static piece of content. This conversational aspect often leads to the question a person is asking being mixed in with other sentences describing the context the person is in, which can be further complicated by single posts potentially containing multiple interlinked questions. The context of a question is generally important to the people who answer the questions, but the conversational nature of online forums tends to produces many answers to a single question. The answer posts are also not straight forward to interpret as they too contain context to frame the answer. The answer posts may still only contain partial answers or only answers to a few of the questions posed in the original question post, while other answer posts may go into unnecessary detailed answers for the information need of the question. The combination of these factors is what makes this a difficult problem in information retrieval, however, there are attempts to solve the problem of effective search through online discussion forums.

An approach used by Cong et al. [1] to analyse forum data is to link questions and answers from threads to improve search performance. They used a two part approach consisting of a pattern based classification method to detect questions in forum threads, along with a graph propagation approach to finding answers in the thread.

This work was continued by Hong et al. [2] who looked at other methods to identify question-answer pairs. Their methods differ from the work by Cong et al. through trying to detect answers without analysing the potential answer post's content. They were able to improve on the methods used by Cong et al. for a question detection method by performing analysis on the question posts rather than by using a pure classification approach. Hong et al. also found that using information about the users who made the posts improved relevance-based retrieval performance. Their ranking scheme that combines authorship, which considers the quality of other posts made by the user, and the position of the answer post, significantly out-performed the other ranking methods they examined.

Another related work is that of Elsas et al.[3] who looked at several algorithms for searching threaded message structures. They found that recognising the message thread structure and that evaluating messages in a thread individually performs better than methods that treat the entire thread as a single

document. This result is validated by the work by Seo et al. [4] who also found that considering the thread structure of discussion forums improves search performance. While their paper focuses on their approach to thread structure identification and annotation, their findings also support the idea that the use of the thread structure improves search performance.

While most prior work was aimed at question answering in general-purpose online forums, this study is specifically about students who are learning programming.

## 4 Methodology

### 4.1 Overview

In order to answer the research questions, a proof of concept system was first developed. This was to confirm how users would interact with a real system and demonstrate the core functionality. The results of the different algorithmic approaches were then compared in terms of standard information retrieval metrics.

#### Experimental Design

As the aim of this research is to compare the different methods to rank results, any information retrieval toolkit would suffice for indexing and querying purposes. Apache Solr[1] was used as it supports query time result boosting.

The 2013 first year discussion forum data was used as the test collection to be searched. The first step was to create a new document collection within Solr, into which the data could be indexed. To this collection a list of stopwords (common words that can be ignored, such as 'the') was added, along with a few other configuration settings that were based on the example simple configuration settings in the Solr documentation [6]. The data was converted from an SQL database dump (obtained from the learning management system) into an XML formatted file that conformed to the Solr XML specification for adding documents to a collection. This XML file was then added to the Solr collection and indexed through the POST tool provided with Solr.

To answer the research questions, 2 different methods for ranking results were compared. The domain-specific method used Solr's query time boosting to change the order of the results; this is a mechanism to multiply the effect of one or more fields when calculating a score for each result, thereby boosting the results (moving them higher up the result list) if they contain higher values in those fields.

The first ranking method (NORM) was to take the results as Solr returned them. The second method (AUTH) was to boost results that were made by authoritative figures; in this context these would be lecturers and tutors.

In Solr the process of performing query time boosting is done through user specified query parameters. This includes which query parser to use, which fields

[1] http://lucene.apache.org/solr/

to boost and which fields to search. For this experiment the dismax query parser was used as it is designed for working with simple queries and supports field boosting [7]. The query boosting was achieved through the boost query parameter, which boosts all results based on the searchable fields listed in the parameter. For the AUTH system the boost query parameter was set to the lecturer and tutor fields; these are boolean fields indicating whether or not a lecturer or tutor posted a particular message.

### Test Query selection

To compare the ranking methods, a set of queries is required. This set of queries was randomly drawn from the 2014 discussion forum data. This was done to provide a more realistic set of queries as, in practice, the system would be exposed to student questions that are expressing an information need, in the form of a discussion forum post.

The process for randomly selecting queries was as follows. All the top level comments in the 2014 data were randomly sorted and then the first 32 questions were selected. The questions were selected at the researchers discretion as to which top level comments constituted a post containing a question. This selection process was used as not all top level posts are questions or, more specifically, the criteria for selection was that the post expressed an information need.

### User Study

Once the list of queries was established, each query was submitted to Solr, through the Solr query endpoint Web service. Each query was submitted twice, once for each ranking method. The number of results for a single query on each method was restricted to only the top 30 results. This cut-off was also chosen as the final system will only display a very small number, much fewer than 30, of results to the students. The distinct results for each query, across the 2 ranking methods, were then combined to form a superset of unique results for each query. This superset is what was given to participants to assign relevance scores to each result.

The process of how the participants provided feedback was as follows. Each participant was randomly assigned 8 queries (of the 32) and each query was assigned to 5 people at random, thus making a total of 20 participants. Then, for each result to each of the 8 queries the participant was assigned, they rated how relevant that result was to the given query on a scale from 1 to 5. This scale was used to allow for a variable amount of relevance for each result as some results are more useful than others; in addition to this the scale allows for an average relevance for a given result to be calculated.

This study was conducted through an instance of the limesurvey tool hosted by the research group.

### Data Analysis

After all the results had been given a relevance rating, the information was moved out of limesurvey and split up into the results for the 2 ranking methods. To calculate the precision, recall and mean average precision (MAP), a binary classification of relevance is required. In binary relevance, a result is either relevant or not. To do this a cut-off of 3 was chosen; this means that any result with an average relevance of 3 or more was deemed to be relevant. As the relevance of every document in the collection was not known, the whole-set approach for recall and precision could not be used; rather these metrics were calculated at various ranks within the results, e.g., recall for the top 5 results. The significance of using recall and precision at 3 and 5 is that those are typical of the number of results shown to users in a popup search interface.

Recall and precision were first calculated for each of the 30 result subsets, averaged over the users who repeated the assessment for each query. The MAP was then calculated using the precision values. The final measure that was calculated was the normalized discounted cumulative gain (NDCG); this is the only measure that does not require results to be given a binary relevance class and as such was calculated with the average relevance assigned by the participants. All these metrics were implemented in their standard formulations, as outlined earlier.

## 5 Results

The results are given in Figure 1 and Table 1.

Table 1 shows the precision, recall, MAP and NDCG of the different methods at various result ranks. Figure 1 shows the interpolated precision-recall curve for the systems.

| | NORM | AUTH |
|---|---|---|
| P@3 | 0.3021 | 0.3021 |
| R@3 | 0.4732 | 0.4911 |
| P@5 | 0.3 | 0.3 |
| R@5 | 0.5427 | 0.5617 |
| MAP | 0.3793 | 0.3882 |
| NDCG@3 | 0.5955 | 0.6034 |
| NDCG@5 | 0.6139 | 0.6211 |
| NDCG@30 | 0.8292 | 0.834 |

**Table 1.** Table of the precision, recall, MAP and NDCG at various ranks. P@3 denotes the precision at the third result. R@5 denotes the recall at the fifth result. NDCG@3 denotes the NDCG at the third result.

Both ranking methods achieve a recall of nearly half at rank three along with a precision of a third. This means that within the first three results for

the ranking methods, one out of the first three results is relevant and about half the relevant documents for the query are in the first three results. This does not change significantly by the fifth result.

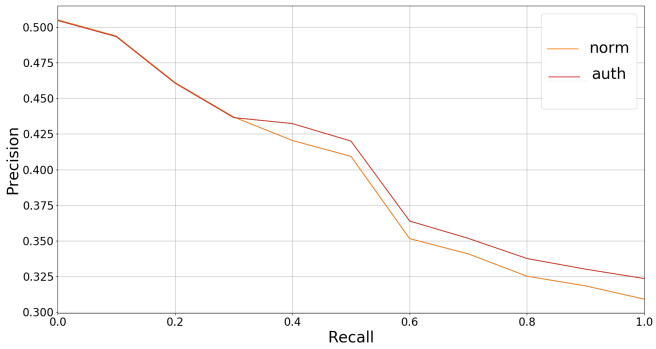


**Fig. 1.** Figure showing the precision-recall curve for the different ranking methods.

Both ranking methods are very similar in terms of overall performance, with the AUTH ranking methods having higher recall for lower ranked results. While the difference is not substantial over the complete result set, the AUTH ranking method produces the highest quality results.

The MAP of the systems is very similar, with the AUTH ranking method performing better than the NORM method. This is a continued trend as seen in all the other metrics. From the above results it is clear that boosting results that are made by authoritative figures yields improvements over a standard information retrieval system.

## 6 Conclusion

This research aimed to determine if information retrieval methods can be applied to give students answers based on previous years' interactions, and if domain-specific features can be exploited for this. Results indicate that indeed it is possible to find relevant answers using information retrieval algorithms, and that the relevance of answers can be improved by exploiting authority as a domain-specific feature. Future work can investigate other domain-specific features in this domain to improve on answerability of questions without human intervention.

## 7 Acknowledgments

This research was partially funded by the National Research Foundation of South Africa (Grant numbers: 85470 and 105862) and University of Cape Town. The

authors acknowledge that opinions, findings and conclusions or recommendations expressed in this publication are that of the authors, and that the NRF accepts no liability whatsoever in this regard.

## References

1. Cong, G., Wang, L., Lin, C.Y., Song, Y.I. and Sun, Y., 2008, July. Finding question-answer pairs from online forums. *In Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 467-474). ACM.
2. Hong, L. and Davison, B.D., 2009, July. A classification-based approach to question answering in discussion boards. *In Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval* (pp. 171-178). ACM.
3. Elsas, J.L. and Carbonell, J.G., 2009, July. It pays to be picky: an evaluation of thread retrieval in online forums. *In Proceedings of the 32nd international ACM SIGIR conference on Research and development in information retrieval* (pp. 714-715). ACM.
4. Seo, J., Croft, W.B. and Smith, D.A., 2009, November. Online community search using thread structure. *In Proceedings of the 18th ACM conference on Information and knowledge management* (pp. 1907-1910). ACM.
5. Balaji, M.S. and Chakrabarti, D., 2010. Student interactions in online discussion forum: Empirical research from 'media richness theory' perspective. *Journal of Interactive Online Learning*, 9(1).
6. Apache Solr Reference Guide. 2018. *Solr Tutorial.* [ONLINE] Available at: https://lucene.apache.org/solr/guide/7_4/solr-tutorial.html. [Accessed 13 November 2018].
7. Apache Solr Reference Guide. 2018. *The DisMax Query Parser.* [ONLINE] Available at: https://lucene.apache.org/solr/guide/7_4/the-dismax-query-parser.html. [Accessed 13 November 2018].
8. Christopher, D.M., Prabhakar, R. and Hinrich, S., 2008. Introduction to information retrieval. An Introduction To Information Retrieval, (pp. 151-175).
9. Baeza-Yates R, Ribeiro-Neto B. 1999. Modern information retrieval. New York: ACM press.
10. Voorhees, E.M. and Harman, D.K. eds., 2005. TREC: Experiment and evaluation in information retrieval (Vol. 63). Cambridge: MIT press.
11. Lupu, M., Salampasis, M. and Hanbury, A., 2014. Domain specific search. In Professional search in the modern world (pp. 96-117). Springer, Cham.
12. Larson, B.E. and Keiper, T.A., 2002. Classroom discussion and threaded electronic discussion: Learning in two arenas.*Contemporary Issues in Technology and Teacher Education*, 2(1), pp.45-62.
13. Singhal, A., 2001. Modern information retrieval: A brief overview. *IEEE Data Eng. Bull.*, 24(4), pp.35-43.
14. Choudhury, N., 2014. World Wide Web and its journey from web 1.0 to web 4.0. i *International Journal of Computer Science and Information Technologies*, 5(6), pp.8096-8100.
15. Suleman, H., 2010. A Social Networking Approach to Integrated Communication in OLEs, Proceedings of South African Computer Lecturers' Association (SACLA) 2010, 3-5 June 2010, Pretoria, South Africa.