Sift through Online Programming Discussions: Effective Search-Result Navigation via Interactive Visualization

Vishal Mehta  
School of Computing, Informatics & Decision Systems Engineering, Arizona State University, 699 S. Mill Ave., Tempe AZ, USA  
vvmehhta1@asu.edu

I-Han Hsiao  
School of Computing, Informatics & Decision Systems Engineering, Arizona State University, 699 S. Mill Ave., Tempe AZ, USA  
Sharon.Hsiao@asu.edu

ABSTRACT
Online programming discussion forums are widely used by programmers for troubleshooting or various problem solving tasks. Large and ever increasing volume of posts on these communities demands more efforts to read and comprehend, thus making it harder to find relevant information. In this paper we designed and studied an interactive network visualization to represent relevant search results for online programming discussion posts. Results showed that users were able to identify relevant items via visual interface compared to traditional ranked lists. Network visualization demonstrated effective search-result navigation support to facilitate users’ tasks and improve query quality for successive queries. Subjective evaluation also revealed that visualizing search results conveys more semantic information in efficient manner and makes searching more effective.

Keywords
Network Visualization; Discussion Forum; Programming; Interactive Visualization; Text Summarization; Exploratory Search.

1. INTRODUCTION
Forums or discussion boards are popular troubleshooting or problem-solving technologies, especially for online learning. They are free, open and fast-growing online communities (homework-help sites, discussion forums for MOOCs courses etc.) that draw massive user-initiated efforts to contribute, to consume and to interact with the content on the site. In the context of programming learning, such free online discussion sites allow programmers and learners to reach out for help so that they can freely discuss programming problems, ranging from general to specific and simple to complex topics. Sites are for instance, stackoverflow1, Dream.In.Code2, Tutorialspoint3, CodeProject4, etc. A 2011-2012 analysis of 28 million course papers submitted to Turnitin5 (2013) revealed that social networking and other user-generated content sites were cited in 23% of the papers written by students in higher education institutions. The same paper also lists online Q&A sites such as Yahoo! Answers and WikiAnswers as second only to Wikipedia among sources used by students (p.6). These sites therefore not only throw open unbounded topics in the form of questions and answers, but are especially attractive for open-ended problem discussions.

The drastic shift in momentum of learning opportunities from traditional learning objects (textbooks, intelligent tutoring systems, worked examples etc.) to community help is becoming prominent but not yet fully supported and comprehended. Among all the discussion forums, search functionality is usually provided for users to filter large volume of discussion posts in an online forum. Some common algorithms are typically deployed to rank the search results, such as PageRank, HITS, or simple keyword matching etc., which present the search results in a form of ranked list. The ranked list organization appears two major problems: 1) it assumes programming learners know how to search; 2) it demands intensive reading-labor to filter the content quality in the large and ever growing corpus, especially inefficient for novices. With these phenomena may end up resulting in expert-oriented communities rather than an open public available technology for all.

We begin to see more and more intelligent interfaces to support general browsing, exploring, searching and adaptation [15-18, 20-22, 6-8]. However, most of these approaches follow traditional hierarchical clustering paradigm, such as Faceted Browsing, Exploratory Search etc., which utilizes the breadcrumb trails to facilitate searches. Such sites typically still rely a lot on query issuers’ efforts (comprehensive query bank and massive reading efforts to identify relevant documents). They are usually designed with several filtering features, such as sorting, voting, badges and other features to filter the content and help readers sift through massive amount of user-generated contents. However, we argue that these filters tend to point out the extremes (i.e. good/bad or recent/obtated) but not represent the overviews, especially the interrelations (i.e. the 5th answer of Q1 and the best answer of Q4 share the same concepts; the best answer of Q1 can be good enough answers to multiple other questions too). Our goal of this project is to investigate an alternative solution to present search results from one of the most fast-growing information retrieval systems, discussion forums. We designed and studied a flat and responsive interactive visualization to navigate search results on large volume of programming discussion posts. We implemented a network interactive visualization that adapts and visualizes users’ search results in a programming discussion forum. We hypothesized that visualizing search results via network

1Stack Overflow (stackoverflow.com) is an online community of question and answer site for professional and enthusiast programmers. It's built and run as part of the Stack Exchange network of Q&A sites.
2Dream.in.code (www.dreamincode.net) is a reputation-based online community for programmers and web developers. Tutorialspoint (TutorialsPoint.com) is a hub that hosts online programming tutorials for learners. CodeProject (www.codeproject.com) is another online community for programmers.  
3Turnitin.com is primarily a plagiarism detection service that analyzes student papers uploaded by subscribing institutions, but their analysis also includes identifying the sources students cite in their papers.
4Turnitin.com is primarily a plagiarism detection service that analyzes student papers uploaded by subscribing institutions, but their analysis also includes identifying the sources students cite in their papers.
visualization (via various abstract visual cues encapsulated in network visualization) will support users’ navigation in search and exploration process in online programming discussion forums.

In this work, we set out to research the effectiveness of search results navigation via an innovative two-dimensional network visualization, with the premise that not-everyone-is-an-expert-searcher. The rest of the paper is organized with the following structure in order to address the core research questions: 1) How does interactive visualization affect search results exploration in large scaled programming discussion forums? 2) Is network visualization an effective visualization in presenting forum posts search results? The next sections present related work, system design, experimental design, and evaluation results. Finally, we summarize the work and discuss limitations and future work.

2. LITERATURE REVIEW

In this article, our aim is to use the visual interface to reduce efforts of reading and grasping in order to find relevant information on online discussion forums. Visualizations have previously been utilized to improve our ability to process large amounts of data. We reviewed three streams of literature related to this project, discussion forum visualizations, visual approaches of recommendation and educational recommenders, and visualizations in the search context.

2.1 Discussion Forum Visualizations

Similar attempts to incorporate visual interface on online discussion forums have been made previously. Heer, Viegas and Wattenberg [1] designed an interface supporting asynchronous collaboration in context of information visualization with underlying assumption that visualization improves ability to process large amounts of data. User study also showed that users were able to identify trends within dataset by discussing, sharing and commenting on visualizations generated for the dataset. This provided us the motivation to include visualization interface to provide higher-level overview of the search results in online discussion forums because forums contain large amount of data and effective visualization can be used to represent this data concisely. Hoque and Carenini [2] presented a novel topic modelling system to extract key points from conversations on online discussion forums. They provided interactive visual interface to revise the model to generate highlights on the fly as per user feedback. This paper presented a novel human-in-the loop topic modelling approach combined with visual interface to support exploration of large conversations. This motivated us to generate the visualization for SearchViz by keeping users in mind and providing them enough flexibility to configure the interface as per their preference by controlling the level of details and position of components.

In their work [3]; Vassileva and Sun studied about impact of visual interface on participation of users in class based online communities. Visual interface shows contributions of users to an online community to encourage social comparison and more participation. Through user study they found that visual interface makes users significantly more proactive in the community by showing increased participation and engagement. This also provided belief that introducing visual interface (SearchViz) to navigate search results on online discussion forums will impact user contribution towards community strongly and positively. In [4] Hsiao and Awasthi discuss about Topic Facet Modeling (TFM) approach to extract forum post semantic for uncovering latent structural topics. TFM looks like a promising approach to automatically generate discourse semantics for large-scale dynamic discussions. The authors also implemented a semantic visual analytics interface in visualizing forum posts semantics [5].

2.2 Visual Approaches of Recommendation and Educational Recommenders

LinkedVis [6], PeerChooser [7], SetFusion [8] are among several other projects that include richer user interface to provide hybrid recommendations. Educational recommenders have been successfully deployed across many disciplines to assist information seeking, exploration, discovery, diversification, enhancing user experiences and engagement [9]. However, they have not yet been fully exploited in the educational sector. Currently, majority of educational recommenders are conceptual designs or prototypes like [13] and [14]. Only a few have been reported with real system usage as in [12]. Secondly, the most common approach in educational visual recommenders is to provide relevant reading objects (new resources, learning partners, or learning object sequences), for instance a suitable next step to learn a new concept or to help solve a problem as suggested in works [13] and [14]. There has been less work focused on dynamic recommendations based on system interactivity.

2.3 Visualization in Search Context

There are several intelligent user interfaces supporting general browsing, exploring, searching and personalization such as PExWEB visualizing content based results similarity [15]; Resultmaps which implements TreeMap to visualize search results [17] which again follow the breadcrumbs trails paradigm to facilitate search. Several researchers made great use of visualization by combining it with web search [16]. They presented an interactive visualization system by combining a number of algorithms to help users analyze and navigate through a collection of web pages. Various case studies carried out revealed that this system was useful for exploration of data. This encouraged us to combine the visualization with online discussion forums to enhance the search result navigation by reducing reading efforts made by users to find relevant content.
The idea of using network visualization for SearchViz by clustering search results for efficient navigation support was also influenced by similar work [18] carried out by Omar, Michael and Recardo. They designed an add-on to combine with traditional information retrieval systems to present and cluster the search results along timelines. In [19] the researchers identify that in exploratory search, there is significant room for improvement in the way search results are returned to the users. Exploratory searches are performed by the users who need to learn, discover and understand complex topics. Hence in such cases presenting web search by enabling users to visualize, manipulate and organize their search results is more effective to serve search purpose of the users. SearchViz aids users in navigating search results on online discussion forums where users may have poorly formed search goals and visualization interface can greatly assist in finding relevant information by summarizing search results and providing efficient navigation to jump to respective results quickly.

3. SEARCHVIZ: INTERACTIVE VISUAL INTERFACE FOR DISCUSSION FORUMS

As you can see in Figure 1; on issuing query ‘implement multiple inheritance’ the visual interface displays search results in form of clusters centered around keywords like class, object, inherit, implement, etc. Each cluster is represented with distinct color and recommended question posts are categorized into appropriate cluster depending on their textual content. Visual interface presents higher level overview of the search results and can be utilized to navigate to appropriate discussion forum posts by clicking on the question circles. Apart from clustering; we also represent each question circle by most frequent keywords appearing in the document. The dataset for this sample discussion forum is generated by downloading content from Stackoverflow 1 website.

We used Apache Lucene library to index the questions on forum according to their plain text as well as program code content to support multiple search features and enable efficient retrieval for various user queries.

3.1 Provided Search Functionalities

3.1.1 Text based Search

In this mode of operation the query string is compared with the textual content of questions in forum posts and related documents are retrieved by searching through text version of index generated by Lucene. Similarity of query to question posts is calculated by using Vector Similarity Model.

3.1.2 Code based Search

Posts on discussion forums may also have code fragments associated with them. Code based search enables users to retrieve questions with specified code fragments. User query string is compared with code of questions to retrieve search results. In this mode; the network visualization for retrieved search results is generated based on text content of questions while ranked list presents semantic code similarity measure. Users can utilize either of these search functionalities to search for relevant content on discussion forum.

3.2 Semantic information conveyed by Visual Interface

3.2.1 Network visualization to provide overview of search results

Visualization shows search results of query in form of clusters centered around keywords. Each cluster is displayed with distinct color and recommended questions are linked to corresponding keywords according to their textual content to generate network visualization. Link of question circle to a keyword within cluster provides higher-level overview about its textual content.

Apart from keyword connection within cluster; each question circle is also represented by other significant words from within its text. User can control the level of detail for question circle representation appropriately via “Cluster Details” button available on left side of the visualization. By default, question circle uses

![Figure 1: Discussion Forum with Interactive Visual Interface to navigate search results](image-url)
single significant word to represent each question. This can be changed to display three significant words per question in order to get deeper insight about each question from the visualization. “Cluster Details” button allows users to adjust such level of details.

3.2.2 Color intensity to represent Query Similarity
In SearchViz, the similarity of a question circle to query is represented via color intensity (Figure 2). Hence question circles with higher color intensity have more $Tf$-$Idf$ similarity to query as compared to ones with low color intensity. This information can be used to make appropriate navigation choices by users.

3.2.3 Radius of question circle to represent number of Answers
Questions on discussion forums may have multiple answers by various users (Figure 3 left). This information is conveyed in visualization through radius of question circle. Question circle with larger radius contains more answers as compared to question circle with smaller radius. This conveys information about length of discussion on a particular question post. Users may be interested in viewing long discussions for some question topics while opting for brief discussions for others. Visual interface provides cues to make such judgments to the users.

3.2.4 Width of links to represent frequency of keywords
Width of link between question circle and keyword is used to represent frequency count of that keyword within text of question (Figure 3 right). Hence wider links have more occurrence of keywords within that question text as compared to question with narrow links.

3.2.5 Using Visualization for navigating through search results
Visual interface provides multiple cues regarding similarity of questions, length of discussions and higher-level overview of question text as well. User can use this set of information to navigate to appropriate question(s) directly by clicking on the question circle in visualization. Corresponding question text will be highlighted directly in the ranked list and scrolled into user view for further inspection. Table 1 demonstrated concrete example of different search results scenarios via SearchViz interface.

### Table 1. Concrete examples of Search Results via SearchViz Interface

<table>
<thead>
<tr>
<th>Color Intensities</th>
<th>Radius</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question represented by ‘interface’ circle has higher $Tf$-$Idf$ similarity than ‘class’ circle.</td>
<td>Question represented by ‘class’ circle has lower $Tf$-$Idf$ similarity as compared to questions to left of this table.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Width of Links</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question circle represented with keyword ‘loop’ has less number of answers.</td>
</tr>
</tbody>
</table>

| Question circle represented with ‘class’ has more frequency word count of keyword ‘method’ within its question text. | Question circle represented with ‘interface’ has less frequency word count of keyword ‘method’ within its question text. |
3.3 Viewing Details through Ranked List
Questions matching user query are represented in ranked list according to decreasing TF-idf similarity measure towards right side of the visual interface. Users can click on ‘View Answer’ button to analyze answers for desired questions. Many times question posts on discussion forums have code fragments associated with them. Such attached code fragments for questions or answers can be viewed by switching to ‘Code’ view from navigation panel above the ranked list. Default view is set to ‘Text’ view for questions and answers.

4. USER STUDY
A user study was conducted to analyze the effectiveness of network based visualization to navigate search results in online discussion forum. Study was designed to simulate situation of a student using online discussion forum for troubleshooting or problem solving various problems related to computer programming.

20 participants were recruited from [removed for blind review], 14 males and 6 females. They were expected to play role of students capable of using online discussion forums for problem solving tasks related to Java programming knowledge. Each task included two programming questions related to unique concept of Java programming. Figure 4 demonstrates user study procedure.

Participants were asked to provide solutions to two tasks by using different versions of online discussion forum as search tool. (1) Version with visual interface to navigate search results, and (2) Version of forum with traditional interface. The order of version for each tasks was randomized for participants. Figure below provides overview of the study procedure.

4.1 Task Description
Table 2 describes details about programming problems included in the user study. The problems tested participant’s knowledge and enabling them to search about concepts like loops, inheritance, constructor and handling exceptions. Each of the problems was assigned difficulty level (E: Easy, M: Moderate, C: Complex).

<table>
<thead>
<tr>
<th>Task - 01: (Time: 20 mins)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How can you implement multiple inheritance in Java? Describe an instance where doing so may not be a good option. (Difficulty: M, Concept: inheritance)</td>
</tr>
<tr>
<td>2. How can constructor affect class inheritance? (Difficulty: M, Concept: constructor, inheritance)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Task – 02: (Time: 20 mins)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Provide few (3-5) examples of problems associated with using iterations. (Difficulty: E, Concept: loops)</td>
</tr>
<tr>
<td>2. Provide reasons to create custom exception class. (Difficulty: C, Concept: exceptions)</td>
</tr>
</tbody>
</table>

Difficulty level for a problem was determined based on concepts covered by the problem and number of relevant answers/posts providing information about the problem. Exception problem, which was categorized as difficult has been discussed in only one forum post while the easier problems are discussed in relatively more posts. Each participant solved these problems using assigned versions of the discussion forum. Recording their behavior enabled us to uniquely identify their behavior and track performance for each problem. This provided data to compare participant’s efficiency and performance while using both versions of discussion forum.

4.2 Study Setup
Control group: Use discussion forum with traditional interface

Experiment group: Use discussion forum with interactive visual interface for search-result navigation.

We utilized simple repeated measure design to meet requirements of randomization of assigned tasks and conditions. Participants were informed about the aim and procedure of user study followed by completing background survey form aimed to get information about their programming experience, sources of learning and troubleshooting for various problems related to programming. 10 minute tutorial on using the visual interface to navigate search results in discussion forums was provided to the participants. This session is very important as it familiarizes participants with interface and also provides learning about interpreting multiple levels of information conveyed by the visualization. While searching on either version of online forum, participants were asked to mark question-answer posts as relevant if it provided them partial or complete information/solution to answer task questions. This exercise was aimed to evaluate search proficiency and provide information about knowledge sources used to provide solutions to task questions.

Finally participants provided subjective evaluations to rate the visual interface on multiple factors like usability, ease of learning, improving search efficiency and satisfaction. They were also interviewed to provide feedback on improving the system and making it more effective and user friendly.
4.3 Participants
This was aimed to get idea about the educational background of student along with familiarity of programming. We were also interested in getting deeper insight into how students learn a new programming language and how do they search for problems related to programming. Hence background survey form also had various questions asking them to identify various sources that helped them learn programming and perform various problem solving tasks related to programming. This information is very valuable as we can integrate these sources of learning into the existing system to make the process of learning more effective and simpler for the students.

Tables below provide brief overview about participant’s basic information, programming knowledge and sources of learning. (Note: Number of participants: 20)

<table>
<thead>
<tr>
<th>Completion Level</th>
<th>Number of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Completed 0-1 courses</td>
<td>6</td>
</tr>
<tr>
<td>Completed 1-3 courses</td>
<td>5</td>
</tr>
<tr>
<td>Completed 4-6 courses</td>
<td>8</td>
</tr>
<tr>
<td>Completed &gt;6 courses</td>
<td>6</td>
</tr>
</tbody>
</table>

Figure 5: Chart displaying programming experience of recruited user study participants

<table>
<thead>
<tr>
<th>Experience Level</th>
<th>Number of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beginner</td>
<td>6</td>
</tr>
<tr>
<td>Intermediate</td>
<td>8</td>
</tr>
<tr>
<td>Proficient</td>
<td>6</td>
</tr>
</tbody>
</table>

Figure 6: Chart displaying Java programming experience of user study participants

As you can see from above table, we recruited participants with approximately equal distributions spanning varied programming experience in form of number of programming classes completed and their self-reported java proficiency.

Below table summarizes various sources of learning identified by participants. (Note: Participants were allowed to report multiple sources of learning and problem solving)

<table>
<thead>
<tr>
<th>Categories</th>
<th>Number of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discussion forums</td>
<td>7</td>
</tr>
<tr>
<td>Online tutorials</td>
<td>12</td>
</tr>
<tr>
<td>Books</td>
<td>15</td>
</tr>
<tr>
<td>Sources of troubleshooting/problem solving</td>
<td>10</td>
</tr>
</tbody>
</table>

Figure 7: Chart displaying sources of learning and problem solving

From above statistics we conclude that apart from reference books on various programming languages, various tutorials and discussion forums available online are also among major learning sources for programmers. It also adds value to our goal of providing efficient navigation through search results on discussion forums.

5. EVALUATION RESULTS
To evaluate the effectiveness of the proposed visual interface for search-result navigation, we analyzed subjects’ system usages, system logs and users’ subjective feedback to assess the quality of the visual interface.

5.1 System Performance Analysis
In our target context, discussion forums, are considered as traditional information retrieval systems, which typically present search results with one-dimensional ranked lists and put the emphasis on relevant items to be placed in the top of the lists in supporting sequential access. SearchViz implements dynamic interactive visualization to support two-dimensional search results in a network view, which not only encourages users to explore flat and responsive representation on large volume of search-results, but also capitalizes network visualization informatics to present relevant document clusters and their relations. To evaluate how successful SearchViz can facilitate navigation, we explore two parameters to measure interface effectiveness, Search Performance and Search Efficiency.

5.1.1 Search performance
According to system logs, we found that users viewed 9.15±6.67 in traditional ranked list interface and 7.40±3.80 in SearchViz. We did not find significant differences on the average views between these two groups. However, we observed that users read much more coherently when used visual interface and behaved sporadically when using traditional ranked lists. This was an important clue that SearchViz may affect users to effectively find relevant information and resulted in cohesive viewing pattern. On the contrary, in the traditional ranked lists group, users’ viewing
patterns were greatly varied, which resulted in possibly some slow and some quick readers or some impatient and some attentive ones.

5.1.2 Search efficiency
Since the search performance was inconclusive by looking at only one parameter (document views), we considered another parameter (time) to evaluate the effects of search efficiency. We computed the amount of time that each user took to identify the first relevant document. We found that users spent significant more time in visual interface ($M=328.61$, $SD=239.20$) than in traditional ranked lists interface ($M=141.72$, $SD=183.80$), $t(17)=2.2997, p=0.022$. This result was originally counter-intuitive to our understanding, which we assumed that visual interface would have required less time. However, the caveat of identifying a relevant document should also take into account the accuracy of whether the item was actually relevant. It led us to look at the quality matrices of users’ searches in examining interface impacts.

5.2 User Performance Analysis
To follow through the clue of visual interface effects, we examined users’ performances by assessing to what extent the system guide users in improving their performances to recognize relevant documents. We used the following matrices.

Precise measures how much of the answers that were accurately marked as relevant. Precision is defined as fraction of marked answers that are actually relevant, according to equation stated below. The ground truth of answer relevancy was collected by two expert judges manually and thoroughly examined the entire selected corpus (Cohen’s $Kappa=0.589$).

$$Precision = \frac{\text{Answers actually relevant}}{\text{Answers marked relevant}}$$

We found that users achieved significant higher precision when used visual interface ($M=0.849$, $SD=0.186$) than traditional ranked lists ($M=0.544$, $SD=0.388$); $t(18)=3.577, p=0.0022$. This demonstrated that users were able to successfully and effectively identify relevant information with the new network visual interface.

![precision](image)

**Figure 8: Precision differences between interfaces**

5.2.1 Query Quality
In this section we discussed the interfaces’ impacts on users’ queries. What were the choices of query terms? Were their query expansions? Were they meaningful? We found that users on average issued 7.35 queries when used visual interface, which was slightly more than traditional one (6.40 averagely). However, it was not significant of the query frequency, neither was the average amount of words were used in the queries. There were 3.23–3.33 on average of words in each query between visual and traditional groups (Table 3). This only means that both groups of users consistently queried similar amount of times quantitatively. To understand the variability of queries, we have to analyze them qualitatively. Thus, we further examined the query’s semantics. We measured the concepts encapsulated in each query by overlapping each query with Java Ontology, which was developed and applied in [24,25] and can be retrieved here.

![Figure 1](image)

**Table 3. Query Statistics Summary.**

<table>
<thead>
<tr>
<th></th>
<th>Average</th>
<th>Traditional</th>
<th>Visual</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query</td>
<td>6.40±4.01</td>
<td>7.35±4.74</td>
<td></td>
</tr>
<tr>
<td>Words</td>
<td>3.23±0.90</td>
<td>3.33±1.08</td>
<td></td>
</tr>
</tbody>
</table>

We observed that there were patterns of **Query Elaboration and Query Concept Elaboration**. **Query Elaboration** counts the subsequent queries for the same question/domain showed an increase in total words in included within them but those words not being conceptual words. For example:

- $q1 = \{\text{for loop}\}$ and $q2 = \{\text{how to write for loop}\}$

**Query Concept Elaboration** considers the subsequent queries for the same question/domain showed an increase in total conceptual words included within them. For instance:

- $q3 = \{\text{problems iteration}\}$ and $q4 = \{\text{problems iteration arraylist}\}$

**Table 3. An example of Query Elaboration and Query Concept Elaboration**, where the participant did not show any concept elaboration in queries issued while using traditional interface. While in case of queries issued via visual interface, we observe concept elaboration in queries $q2, q3$ and $q5, q6$.

**Table 3. Query Concept Elaboration.**

<table>
<thead>
<tr>
<th>Query</th>
<th>Traditional</th>
<th>Visual</th>
</tr>
</thead>
<tbody>
<tr>
<td>$q1$</td>
<td>${\text{for loop}}$</td>
<td>${\text{for loop}}$</td>
</tr>
<tr>
<td>$q2$</td>
<td>${\text{problems associated with iterations}}$</td>
<td>${\text{problems iteration + problems =}}$</td>
</tr>
<tr>
<td>$q3$</td>
<td>${\text{inheritance + problems}}$</td>
<td>${\text{inheritance + problems =}}$</td>
</tr>
<tr>
<td>$q4$</td>
<td>${\text{constructor after inheriting a class}}$</td>
<td>${\text{constructor after inheriting a class}}$</td>
</tr>
<tr>
<td>$q5$</td>
<td>${\text{create a custom exception class}}$</td>
<td>${\text{create a custom exception class}}$</td>
</tr>
<tr>
<td>$q6$</td>
<td>${\text{custom exception class}}$</td>
<td>${\text{custom exception class}}$</td>
</tr>
</tbody>
</table>

Thus, $q1$ and $q2$ appeared no conceptual expansion, but $q3$ and $q4$ did. Table 3 demonstrated a concrete example for one of the subjects, which **Query Concept Elaboration** patterns were displayed in bold fonts. We found that 40% of the queries from subjects who used visual interface improved their query quality for successive queries. This behavior was also observed in students using traditional interface too, but there was only 10% of the over queries. The results showed that with the support of network visualization, users were able to expand the queries by adding more conceptual words rather than simply extending length of query string. The outcome can attribute to the succinct network nodes view in the visualization, which displayed each relevant document summarization in one or three keywords. The content summarization provided users to quickly grasp documents’ relations with key concepts by recognizing only a few “new” terms, instead of traditionally reading through lengthy texts and mentally forming relations the and picking out “new” concepts for the next query. We also found evidence that participants heavily interacted with the visualization. Averagely each user made 32.30 clicks on the network nodes and 17.75

---

1Source of Java Ontology: [http://www.pitt.edu/~paws/ont/java.owl](http://www.pitt.edu/~paws/ont/java.owl)
unique node clicks. These findings demonstrated that interactive network visualization helped to improve users’ query quality and enhance navigation quality.

5.2.2 Task complexity effects
Based on the concepts involved for the given tasks, we categorized them by complexity in three types, Easy, Moderate and Complex. We measured the task’s completion and accuracy, where completion was calculated based on entirety of task-response in covering all aspects of problem and accuracy is the measure of correctness of the response. We found that users outperformed in both parameters accuracy and completion in topic Exception (complex task) when used visual interface. We did not find such patterns in easier topics (Figure 9 & 10). This showed that interactive visual interface was especially helpful for complex problems.

5.3 Subjective Evaluation
This section summarizes the subjective evaluations provided for various criteria by students to visual interface. Participants provided ratings to access usefulness, ease of use and ease of learning of visual interface.

5.3.1 Accessing Usefulness
While accessing usefulness, students stated that visual interface conveys multiple levels of information concisely and effectively. It was also helpful in finding relevant information more effectively as compared to traditional list based view. Interactive visual representation of search results conveys more semantic information as compared to traditional list based view.

5.3.2 Accessing Ease of Use and Learning
Participants reported that they did not take much time in getting familiar to the user interface and found it relatively easy to learn and understand. They reported that interface helped them in finding relative information easily and effectively. Participants also provided feedback on improving the visual interface by using separate shapes to display keywords and question circles to the difference between the two more evident.

6. CONCLUSIONS

6.1 Summary
In this article we identified the need for providing effective ways to find relevant information in fast growing user-generated content of online discussion forums. We proposed and designed a novel interactive visualization interface, which is capable of providing higher-level overview of results along with efficient search-result navigation to address that need.

We then carried out a user study to analyze the usefulness and impact of visual interface on search performance and efficiency of participants. We observed that participants were successfully and effectively able to identify relevant information while using visual
interface as opposed to traditional interface. We also found that using discussion forum with visual interface improved query quality and achieved higher rates of accuracy and completion while solving complex tasks. This is such an encouraging note for the majority fast growing discussion forums. By providing visual interfaces as alternative in scrutinizing search results can enhance users’ performances in complex problem solving.

6.2 Limitations & Future Work
In addition to the findings, there are a few limitations in current setup. Current network visualization only considers content semantics ignoring users’ connections, which can be a potential alternative multimodal network visualization in enhancing current work. Another improvement possibility is to enhance network features, such as sort the network as hive plots etc. which can potentially provide more detailed analyses and views. In spite of limitations, study analysis and subjective feedback led us to believe that network based visualization interface was effective in navigating search results in online discussion forums by eliminating efforts for extensive reading and grasping capabilities to find relevant information. This novel approach of improving search precision and efficiency in online discussion forums by integrating visual interface with traditional list based result-view should be explored further.

7. REFERENCES


