2017

An Empirical Investigation of Collaborative Web Search Tool on Novice's Query Behavior

Mareh Fakhir Al-Sammarraie
University of North Florida

AN EMPIRICAL INVESTIGATION OF COLLABORATIVE WEB SEARCH TOOL ON NOVICE'S QUERY BEHAVIOR

by

Mareh F. Al-Sammarraie

A thesis submitted to the
School of Computing
in partial fulfillment of the requirements for the degree of

Master of Science in Computing and Information Sciences

UNIVERSITY OF NORTH FLORIDA
SCHOOL OF COMPUTING

August, 2017

The thesis "An Empirical Investigation of Collaborative Web Search Tool on Novice's Query Behavior" submitted by Mareh F. Al-Sammarraie in partial fulfillment of the requirements for the degree of Master of Science in Computing and Information Sciences has been

Approved by the thesis committee: Date

Dr. Karthikeyan Umapathy
Thesis Advisor and Committee Chairperson

Dr. Lakshmi Goel

Dr. Sandeep Reddivari

Accepted for the School of Computing:

Dr. Sherif Elfayoumy
Director of the School of Computing

Accepted for the College of Computing, Engineering, and Construction:

Dr. Mark Tumeo
Dean of the College

Accepted for the University:

Dr. John Kantner
Dean of the Graduate School

CONTENTS

EQUATIONS

FIGURES

TABLES

ABSTRACT

In the past decade, research efforts dedicated to studying the process of collaborative web search have been on the rise. Yet, a limited number of studies have examined the impact of collaborative information search processes on novices' query behaviors. Studying and analyzing factors that influence web search behaviors, specifically users' patterns of queries when using collaborative search systems can help with making query suggestions for group users. Improvements in user query behaviors and system query suggestions help in reducing search time and increasing query success rates for novices.

This thesis investigates the influence of collaboration between experts and novices as well as the use of a collaborative web search tool on novices' query behavior. We used SearchTeam as our collaborative search tool. This empirical study involves four collaborative team conditions: SearchTeam and expert-novice team, SearchTeam and novice-novice team, traditional and expert-novice team, and traditional and novice-novice team. We analyzed participants' query behavior in two dimensions: quantitatively (e.g. the query success rate), and qualitatively (e.g. the query reformulation patterns).

The findings of this study reveal that the successful query rate is higher in expert-novice collaborative teams, who used the collaborative search tools. Participants in expert-novice collaborative teams who used the collaborative search tools, required less time to

finalize all tasks compared to expert-novice collaborative teams, who used the traditional search tools. Self-issued queries and chat logs were major sources of terms that novice participants in expert-novice collaborative teams who used the collaborative search tools used. Novices as part of expert-novice pairs who used the collaborative search tools, employed New and Specialization more often as query reformulation patterns.

The results of this study contribute to the literature by providing detailed investigation regarding the influence of utilizing collaborative search tool (SearchTeam) in the context of software troubleshooting and development. This study highlights the possible collaborative information seeking (CIS) activities that may occur among software developers' interns and their mentors. Furthermore, our study reveals that there are specific features, such as awareness and built-in instant messaging (IM), offered by SearchTeam that can promote the CIS activities among participants and help increase novices' query success rates. Finally, we believe the use of CIS tools, designed to support collaborative search actions in big software development companies, has the potential to improve the overall novices' query behavior and search strategies.

Chapter 1

INTRODUCTION

Information Retrieval (IR) is concerned with retrieving relevant documents from a data source. With the exponential growth of information sources on the Internet, web information retrieval systems have evolved from a simple search tool to an entity that influences many aspects of our lives (Al-Sammarraie & Umapathy, 2016). Web search is a crucial information retrieval problem, and many scientific studies have been conducted to address this topic from the individual user search task perspective (Lewandowski, 2012). However, there are many situations involving a group of Web users working together on complex tasks. Parts of these complex tasks go beyond the simple individual web search process. For instance, a part of the complex project may revolve around searching the Web for relevant information, sharing information among a group of users, comparing and synthesizing various pieces of information from multiple sources, making decisions, and using the synthesized solution(s) (Shah, 2014). There are many day-to-day life examples on the above-presented situations, such as planning a vacation, coauthors who are working together on a scholarly article, an engaged couple who are planning a wedding, and a recruitment team that is working on a new hiring project (Shah, 2014). All these are examples of collaborative information seeking (CIS) projects. While these projects go beyond the simple search process, they also have a common goal that must be mutually beneficial to all participants.

Important contributions have been made from prior research works to explain the relationship between CIS and other fields of information seeking, collaboration, and information retrieval. The paper by Shah, Capra, & Hansen (2014) defined CIS as it is not only an information-seeking activity including collaboration, but also a collaborative activity with the goal of information seeking. Earlier research efforts have also identified several important aspects of the CIS process (Shah, 2014). First, there should be a common goal that brings participants together for collaboration. Second, the CIS process should be initiated to help solve complex tasks there are few benefits from collaborating on simple fact-finding information tasks. Third, collaborations are only useful if the processing load or overhead is acceptable for a given situation. Lastly, the CIS process should take place among a small group of participants with different levels of skills and roles.

There have been ongoing efforts to produce systems that would connect multiple users to facilitate collaborative information seeking, either by creating new IR systems or by extending existing IR systems to accommodate more than one user (Shah, 2014). The functionalities offered by each system are different and can vary based on the context of use. However, the common goal shared by these systems has been to reduce the need for workarounds by offering browser-based support for collaborative search (Kelly & Payne, 2014). Awareness, division of labor, persistence, and sensemaking are among the aspects of the collaborative search that most of the collaborative search tools aim to support (an explanation of each aspect is provided in Chapter 2).

While previous research studies have advised that the features mentioned above are useful for supporting collaborative information seeking, in general, further investigations are needed to study the impact of these tools and their features in other domain specific settings, such as in the context of software development. Studying the impact of these tools in the domain of software development can help provide us with insightful details for how software programmers collaborate to troubleshoot issues and what type of tools they need to make their collaboration more effective.

1.1 Problem Statement

Internship programs vary widely among software development companies, but usually all these programs assign interns to a project and pair them with another engineer who acts as their mentor. In small companies, it might be the case that the intern doesn't have a specific mentor, but has a person or group of individuals he or she can go to for help.

Typically, interns work with their mentors either on the development of a new software feature or on maintaining an existing one. In either case, searching the Web for a solution that aims to solve a particular problem is one of the responsibilities that an intern is expected to perform on a daily basis. This task is sometimes challenging for new interns, since they lack the knowledge and the practical skills required. One of the issues interns might face is that they spend too much time trying to find the correct terms or keywords to use in their search queries to locate the useful content. In addition to this, they sometimes struggle with finding reliable sources that offer a trusted solution to their problem. Moreover, interns might face issues with implementing tasks involving analysis

and decision-making for a list of search results they have retrieved during a search session. All this can often lead to problems completing the work and, thereby, delaying the task.

Interns and mentors are expected to solve problems collaboratively so that interns can learn best practices, such as relevant keywords to use and tricks to identify relevant sources. Most of the time, mentors may use ad hoc practices, such as sharing links via email or IM and using blogs or text documents to keep track of search results, while troubleshooting an application's problem collaboratively with an intern. While this approach seems to provide a solution for part of the problem for a short time, it becomes arduous and time-consuming. The interaction between interns and their mentors, such as in requirements clarification process, Web searching tasks, search results analyzing, and the decision making process, is basically a CIS process. For interns and mentors to perform their CIS activities effectively, they need simple and friendly collaborative Web search tools.

The goal of this research is to explore the impact of Web users' domain expertise on their query behavior in a collaborative web search using collaborative search systems. Furthermore, it aims to identify whether the CIS process has any positive impact regarding augmenting querying strategies for novice users when paired with domain experts (Al-Sammarraie & Umapathy, 2016). We believe that the results of this study will be beneficial for directing future design activities of collaborative search tools.

1.2 Contributions

The majority of prior research studies reported in the literature have investigated the benefits of collaborative search tools when used in supporting everyday tasks, such as travel planning and online shopping. This thesis, however, is conducted to study the impact of using these collaborative search tools in a domain specific setting. Therefore, the results of the study will expect to add knowledge to the field by providing insightful details for how software programmers collaborate and what type of tools they need to make their collaboration more effective.

Additionally, the results of this study clarify the impact of using collaborative search tools on knowledge-sharing among team members. In particular, we investigated the impact of such tools on novice users' querying strategies when paired with a domain expert.

Finally, this study tested the hypothesis that having domain experts work with non-experts using collaborative search systems will immensely increase the query success rates for non-expert users and help them learn better querying strategies over the course of time. Proving the above hypothesis will suggest that it is highly recommended to use collaborative web search tools during the training of interns.

1.3 Plan

We used two existing Web search tools: SearchTeam and Google Integrated Platform. SearchTeam (http://searchteam.com/) is a free collaborative Web search system that can be used by a single user or a group of users. Google Integrated Platform includes Google search engine, Google Docs, and Google Hangout. We designed a laboratory study of CIS tool use with the purpose of acquiring a better understanding of the difference in collaborators' search behavior when using a collaborative Web search tool as opposed to traditional search tools like Google. This study also aims to investigate the impact of such tools on novice users' querying strategies when paired with a domain expert.

The experiment was conducted on a group of users who are expected to collaborate on a list of search tasks relevant to troubleshooting ASP.NET MVC codebase. The participant population consisted of two groups. The first group included University of North Florida (UNF) School of Computing students, and the second group includes senior software developers from local companies. The participants were classified as novice users and expert users, and then organized into teams of two. Each team was randomly assigned to use either a collaborative search tool (SearchTeam) or a traditional search tool (Google). The collaborative team type conditions were organized into four groups.

The variables of the study are query success rates, time spent to complete the tasks, query patterns, and participants' perspective on the sources that affected the way they issue

search entries. Data on participants' perspective were collected from transaction and chat logs as well as via an online questionnaire. The questionnaire contained questions on participants' experiences using the collaborative search tool.

1.4 Organization

This thesis is divided into six chapters the current chapter provides an introduction. The second chapter has an overview of the collaborative information seeking (CIS) concepts, and the classification of query reformulation activities. It also includes a literature review, which analyzed the current state of the art research in CIS and query reformulation behavior. A summary of previous efforts relate to the goal of this thesis is provided. The focus has primarily been on papers that offer insight on how to analyze the impact of CIS on Web users' actions during a Web search. The third chapter has included the research method and the design of experiments. The collaborative team type conditions and the experimental procedure are presented. The fourth chapter has included the data collection process and the measures conducted to evaluate the impact of CIS on users' query behavior. The fifth chapter provided conclusion on the state of this research, and its anticipated findings and potential contributions. In the sixth chapter, a conclusion for this research study is presented.

Chapter 2

BACKGROUND AND LITERATURE REVIEW

2.1 Background

In this chapter, an explanation of various concepts relevant to this empirical study is provided. The two main concepts are collaborative information seeking (CIS) process and query behaviors in Web search. These two topics have been discussed at length to provide an overview for understanding the study's key elements. A brief overview of other relevant areas used in this research, which includes the level of expertise, user roles, and frameworks for CIS, has also been provided. Understanding the concepts specified above is necessary for designing our experimental study.

2.1.1 Collaborative Information Seeking (CIS) Approach

Web users often work together when planning a complex search task, exploring the different information sources, collecting domain-related data, and making sense of this data (Stange & Nurnberger, 2015). However, traditional information seeking tools do not support this behavior, and this lack of adequate tools causes people to adopt ad hoc workarounds such as sharing links via email or IM or over-the-shoulder surfing (Golovchinsky, Qvarfordt, & Pickens, 2009). In addition to that, the ever-increasing number of internet-connected devices and the growing need for online sharing and collaboration, have led several research groups to explore different aspects of

collaboration in support of information seeking and have recognized the potential benefits of collaborative search tools (Shah et al., 2014).

Collaborative information seeking (CIS) resembles individual information seeking approaches, but with added dimensions such as the roles that collaborators assume, how collaborators work together across time and space, the awareness of another collaborator's actions, and the negotiation and knowledge sharing that must happen (Shah et al., 2014). For example, collaborators might take peer or asymmetrical roles, such as research/reader or student/mentor. Additionally, they might involve CIS activities that can occur synchronously or asynchronously, be co-located or remote, or require specialized search systems (Shah et al., 2014).

The following elements need to be present when performing a collaborative information search to accomplish successful collaboration among a group of people as Shah emphasized in his study (Shah, 2014).

1) *Common goal and mutual benefits*. The process of collaboration includes people working together for a common goal and can help produce something that is more than the sum of individual participants' contribution. As Shah explained in his study (Shah, 2014), the procedure of collaboration is not provided as an obvious functionality by most of the collaborative information systems. Rather, these systems are designed in a way to provide support for people who want to work collaboratively toward a common goal.

2) *Difficult and complex search task.* There will be no benefits from working collaboratively on simple search tasks such as fact-finding tasks. (Shah, 2014) claimed

there is no point in collaborating if the task is simple; however, collaboration can occur when multiple individuals are working on a task that is exploratory in nature.

3) *Ratio of high benefits to overhead*. Collaboration can be successful if it meets or exceeds the benefits expected for it. The simple strategy of divide-and-conquer may have overhead. To achieve collaboration, such overhead should be within the acceptable limit.

4) *Lack of knowledge or skills*. One of the common reasons to seek collaboration is the lack of knowledge or skills individuals have to solve a particular problem. Usually, people will tend to collaborate if they lean toward achieving better results than what they each could do individually.

At the different levels of the information seeking process, collaboration among a group of users can occur. For example, collaboration can happen while users are formulating an information request, gaining search results, or while organizing and using search results (Shah, 2010). Moreover, users participating in collaborative search may either take the same roles or different roles in the search task based on the relationships among them. For example, in a symmetric collaboration, the collaborators share a common information need and fulfill the same roles in the search, whereas in an asymmetric collaboration, the participants fulfill different roles in the search task (Morris & Teevan, 2010). The fulfillment of different roles in asymmetric collaboration search may take place as a result of dividing a search task into roles based on the nature of the task, an individual's familiarity with technology, or specific expertise (Morris & Teevan, 2010). In this study, we are interested in studying the effects of collaborators' level of expertise on their search behavior and the roles they fulfill to accomplish a search task.

2.1.2 Aspects Supported by Collaborative Information Seeking Tools

The main principles encompassed in the collaborative information retrieval (CIR) process that collaborative tools seek to support are: the division of labor, awareness, and knowledge sharing (Soulier, Tamine, & Bahsoun, 2014). The *division of labor* allows collaborators to divide the workload among them. Furthermore, the division of labor is achieved by following one of two approaches. The first is a task-based approach, assigns different search tasks among collaborators. The second is a document-based approach which splits search results in order to display to each participant different lists of documents (Morris, Teevan, & Bush, 2008). (Foley & Smeaton, 2010) in their study also outlined several ways in which the division of labor can be achieved. The authors refer to three strategies that are used in collaborative writing namely: *parallel*, where the task is divided into subtasks and collaborators work simultaneously; *sequential*, where the work is divided into several stages and the output from one stage is handed to next user; *reciprocal*, where collaborators work together and mutually adjust their activities. The same authors also linked the achievement of the division of labor to user roles, where each group member is assigned certain jobs according to their role. These user roles are usually defined based on the skills and expertise of individuals within an organization. *Awareness* is like an alert for users to prevent duplicate work, such as notifying users of already visited links, already seen documents, or previously submitted queries. Collaborative browsers support the awareness principle by providing a shared workspace, allowing collaborative users to be aware of each other visited and saved web sites' links, and to learn from them (Soulier et al., 2014). *Knowledge sharing* allows the exchange of information among users by using shared workspaces or adapted tools supporting

brainstorming among collaborators, such as the use of instant messaging (Soulier et al., 2014).

In addition to the three aspects presented above, the study by (Kelly & Payne, 2014) has also included persistence and sensemaking as general aspects that most collaborative search tools support. Persistence refers to the storage and display of activities that are made in prior search sessions. Persistence has been supported through the retention of chat logs, pageview statistics, and session histories. While persistence supports asynchronous collaboration through re-finding of information and resumption of previous search sessions; sensemaking supports the understanding of the search process in terms of what has been found, how it was found, and where tasks have been assigned between collaborators (Kelly & Payne, 2014). For example, sensemaking can be supported through enabling of access to timelines of pages viewed by collaborators, and the visualization of search strategies and trajectories (Kelly & Payne, 2014).

2.1.3 Classification of CIS Frameworks

CIS systems, in general, can be classified along four dimensions: depth of mediation, intent, time, and location (Golovchinsky et al., 2009).

- Depth of Mediation (System Mediated vs. User/Interface-Mediated Collaboration)
    - In collaborative projects, collaboration can be completely driven by the system through an underlying algorithm. In this case the system acts as an active component for collaboration. There are two examples of this system. The first is a collaborative filtering system which is designed to keep track of

each user's data separately, before combining it to make specific recommendations (Shah, 2014). The second example is a system-mediated CIS system which is similar to the first with one difference the user's intention to collaborate (users of system-mediated CIS tools have explicit involvement in the process).

    - Alternatively, collaboration can be completely driven and controlled by the collaborators with support from a systems' user interface. The system search engine (the back-end) is unaware of the contributions of different people such as issued queries and saved documents. Therefore, the system serves as a passive element helping with aspects such as communication and awareness. Examples are SearchTeam, Coagmento and Ariadne. Users of these systems use the co-browsing interface provided by the system to work through their information seeking process. However, the system is not doing anything more than responding to users' actions.

- Intent (Explicit vs. Implicit Collaboration)
    - This dimension describes how explicitly collaboration is defined.
    - It is used to distinguish truly collaborative systems from the collaborative filtering systems that infer or provide recommendation based on other users' actions or behavior.
- Concurrency (Time): Synchronous vs. Asynchronous
    - This dimension indicates whether the collaborative activities between collaborators are concurrent.

    - Synchronous collaboration occurs when people influence each other's searches in real time.
    - Asynchronous collaboration occurs when prior searches, either personal or aggregated from a community, impact later searches (Golovchinsky et al., 2009).
    - CIS can support both situations. An example is when users of SearchTogether save search results for people who might not be online (Asynchronous CIS), or they can issue a search query and go over the results at the same time (Synchronous CIS).
- Location: Co-located vs. Remote
    - Collaboration may happen among a group of people who are working in the same place at the same time or it might.
    - Occur among a group of collaborators who might be distributed.
    - The remote circumstance, in which collaborators are distributed, increases opportunities for collaboration but decreases the prospects of communication (Golovchinsky et al., 2009).

2.1.4 Roles in Collaborative Information Seeking Systems

Collaborative models assume that people are different and by collaborating they can utilize other searchers' expertise, skills, or search strategies to solve a shared information need. Within a collaborative session one collaborative individual can have more knowledge about the search task or topic compared to the other collaborators (Soulier et al., 2014). Additionally, within a search session, collaborators can be characterized by the

difference of domain expertise level, and they can be assigned to distinct roles to work towards a shared information need (Soulier et al., 2014). The configuration of users' roles might be implicit in the functionality of the interface for example, specifying queries and making relevance judgments. Or they can be more explicit, such as when people use different interfaces for different subtasks (Golovchinsky et al., 2009). Previous research efforts surrounding roles classification and support have identified different pairs of roles collaborators can take and can be categorized into two groups.

The first group is focused on different combinations of the searchers' expertise (expertise-based roles), namely peer, domain A expert/domain B expert, search expert/domain expert, and domain expert/domain novice pair. The second group is focused on searchers' activities during the search, namely prospector-miner pair (Golovchinsky et al., 2009) (see Table 1).

Roles Configurations		
Roles focus on different combination of the searchers' expertise	Peer	• All Participants use the same interfaces to control the system and coordinate their activities. • Participants use their systems independently and combine their results manually.
	Domain A expert/ domain B expert	• Participants use symmetrical interfaces. • Collaboration occurs between people with different domain knowledge.
	Search expert/search novice	• Collaboration occurs between people with different level of familiarity with search tools.
	Domain expert/domain novice	• Collaboration occurs between people with different level of expertise with a domain knowledge.
	Search expert/domain expert	• Collaboration occurs between a skilled searcher and a person with a complex information need.
Roles focus on searchers' activities during the search	Prospector/miner	• In these roles, one participant searches broadly and the other search deeply. • The prospector issues many queries with minor relevance judgments for each search result. • The miner makes detailed relevance judgements on results found by the prospector. • It is important to note that the specialization role is driven by the division of the search task into subtasks and not by user's knowledge, such as in search expert/domain expert pair.

Table 1: User Roles in CIS Systems

2.1.5 Query Reformulation

Web information retrieval process can be divided into three stages: finding documents, query formulations, and determining document relevance (Al-Sammarraie & Umapathy, 2016). Prior research studies have mentioned that the successful assistance for query-reformulation must be designed based on the understanding of users' query behaviors (Mohammad Arif, Du, & Lee, 2014; Yue, Han, & He, 2013). Therefore, in this study we focus on the second stage, query formulations, as it requires user action. In particular, we study query formulations from the collaborative search task perspective.

Web searching is a crucial information retrieval problem and was defined by (Mohammad Arif et al., 2014) as the process of querying and reformulating queries to fulfill information needs. Complex and explorative Web searches often involve frequent modifications of Web searchers' queries to obtain better results, and this process is referred as query reformulation (Jansen, Booth, & Spink, 2009; Mohammad Arif et al., 2014). The query formulation and reformulation process is an important topic, not only in individual searches, but also in collaborative searches, which covers issues such as patterns of query reformulation and the reliable sources for query expansion (Mohammad Arif et al., 2014). Prior research pointed out that knowing how and when groups of users issue queries and reformulate queries during the collaborative search process makes the targeted query suggestions offered by the system more efficient and effective in supporting the collaborative process (Yue et al., 2013).

Different classifications of query-reformulation types have been identified by the literature. The authors (Jansen et al., 2009) in their study have presented six states of query reformulation. Each state is defined as following:

Query State	Description
New	Is this first query from a user, or a query on a new topic from a searcher?
Assistance	The query is generated by a user who has used a "Are you Looking For?" feature provided by the system, such as the feature "Did You Mean?" offered by Google, which focuses on spellchecking.
Content Change	The user issued, or executed, a query using another content collection.
Generalization	The current query is on the same topic as the previous query issued by a certain searcher, but he or she is now seeking more general information. The state of generalization is determined when the query has fewer terms than the previous query by a certain searcher.
Reformulation	The current query is on the same topic as the previous query issued by a certain searcher, but both queries have common terms. The state of reformulation is determined when the query has the same number of terms as the previous query by a certain searcher with at least one term being in both queries.
Specialization	The current query is on the same topic as the previous query issued by a certain searcher, but he or she is now seeking more specific information. The state of specialization is determined when the query has more terms than the previous query by a certain searcher.

Table 2: Classification of Query Reformulation

2.2 Related Work

While there has been a growing interest in understanding, and supporting CIS practices, we still lack clear understanding of people's CIS behavior (Al-Sammarraie & Umapathy, 2016) . Individuals involved in CIS would be understanding search tasks, formulating queries, determining results relevance, sharing results, and using information collaboratively with their partners (Paul & Reddy, 2010). CIS activities are supported by collaborative search tools. Thus, investigating and gaining better understanding of these

activities is important for designing and developing effective collaborative web search tools (Al-Sammarraie & Umapathy, 2016). Collaborative search tools are expected to provide awareness features that archive the groups 'query histories, selected results, and comments, as well as division of labor features, which include chat systems, ability to divide search tasks, and selecting search results based on a group member's action (Paul & Morris, 2009). Collaborative search tools have potential to be used by business colleagues to find information related to their work, and by friends and family members for vacation planning or seeking medical information (Morris, Lombardo, & Wigdor, 2010). Below, we review literature relevant to query reformulation in the CIS context.

2.2.1 Query Reformulation Patterns in Collaborative Web Search

Individual users reformulate their queries in an iterative process during web searches until they successfully retrieve the most relevant results. Similarly, users in collaborative web searches have to modify their initial queries to reach the expected results. There have been few research efforts investigating different classifications of query reformation in the context of collaborative web searches. Studying different types of query reformulation in collaborative web searches can provide valuable information about the interactions between a group of users and web search systems. The acquisition of such knowledge will not only benefit the collaborative web search systems, it will also support and empower query suggestion features in web search engines utilized by individual users as well (Al-Sammarraie & Umapathy, 2016).

A study conducted by Arif et al. shows five different operations performed during collaborative query reformulation to finish a tourism-related exploratory web search task (Mohammad Arif et al., 2014). Addition was the most used operation (61%), followed by Modification (20%), Addition and Reorder (14%), Reorder (4%), and Addition and Modification (1%). Yue et al. (Yue et al., 2013) defined four types of query reformulation: New, Specialization, Reconstruction, and Generalization. Two exploratory web search tasks were used in the study. The authors compared types of query reformulations done by individual users with those done by a team. The statistical results show that participants in collaborative searches tended to use the New and Specialization patterns, while those in individual searches were more likely to use Reconstruction.

2.2.2 Factors Influencing Query Reformulation in Collaborative Web Search

Few studies have been conducted in regard to factors influencing query reformulation in the context of explicit collaborative information seeking (CIS). The Study by Yue, Han, He, & Jiang (2014) conducted experiments to examine factors influencing search term reformulation with a collaborative search process. They developed a collaborative web search system (CollabSearch) to conduct the study. They assigned two types of search tasks: the first was academic, a recall-oriented and information-gathering task; while the second was leisure, a utility-based, decision-making task. Results of their study revealed three kinds of influences. The first category is query reformulation based on search-related action. For example, participants' previous search histories are considered a very reliable source for relevance feedback and can consistently improve search results. The second grouping is collaborative action-based query reformulation. In this second group,

the authors specified three different types of collaborative actions that have influenced the creation of new query terms: a user checking his or her partner's saved documents; a user checking the partner's query history, and explicit communication between the partners through text messaging/chat. The third and last influence is using chat. The authors concluded that chat provides overall guidelines in search for academic tasks and specific search topics in leisure-oriented tasks.

Another study conducted by Yue et al. (Yue et al., 2013) investigated the effect of collaboration and task type on users' query reformulation behavior. They compared the results of three aspects: query features, query reformulation behaviors, and performance for participants working under two conditions: collaborative search and individual search. The findings revealed that both collaboration and task type affect users' behaviors in exploratory web search. Studies also show that various factors, such as task complexity and users' domain expertise, influence web search behaviors in general (Monchaux, Amadieu, Chevalier, & Mariné, 2015). Prior research has also shown the influence of involving experts within the CIS process. Kang et al. in their study compared how experts and novices performed exploratory search using a traditional search engine and a social tagging system (Kang, Fu, & Kannampallil, 2010) . They also found that experts were better at finding information using both interfaces, because experts were better at interpreting social tags in the tagging system and generating search keywords.
In summary, previous research has found that differences in domain expertise may impact search behavior. Yet, little is known about the impact of the CIS process on novices' query behaviors when paired with domain experts. Therefore, it is worth

investigating how pairing novices and experts influences query formulation behavior of novice users when using a collaborative search tool (Al-Sammarraie & Umapathy, 2016).

Chapter 3

RESEARCH METHODOLOGY

In this chapter, the experiment design elements and the decisions made during each step of the design process are discussed. The research method and the variables included in the study are explained first. An overview of the collaborative tool that is used by participants during the experiment is presented. The recruitment plan and the different collaborative team type conditions are also clarified. A justification of the selection of the search tasks that are used to conduct the search sessions is then presented. Finally, the experiment procedure that the investigators followed to collect the search data is provided.

3.1 Research Question and Hypotheses

This study addresses the following research questions:

RQ: Are there any differences in the query success rate, task completion time, query sources of terms, and patterns of query reformulations caused by collaborative teams (expert-novice teams) using either a traditional search engine (Google) or a collaborative web search tool (SearchTeam)?

The first hypothesis was that collaborative teams, who had an expert member and used CIS tools, would:

- Have higher success rates for search tasks.
- Require less time to finish search tasks and achieve higher query performance rates.

The second hypothesis was that novices in collaborative teams, who worked with an expert member and used CIS tools, would

- Incorporate self-issued terms and keywords provided by expert member in their successful queries.
- Use New and Specialization reformulation query types more often.

Finding the results of the two measures (query success rate and task completion time) presented in the research question, will help assess the effectiveness of CIS tools as opposed to traditional web search tools in the context of novice-expert pairs troubleshooting software development problems. Finally, by analyzing results of the additional two measures (query sources of terms, and patterns of query reformulations) presented in the research question, will help knowing about the impact of using such tools on novices' query behavior and search strategies.

3.2 Experimental Design

The purpose of this study is to investigate the impact of pairing novices with experts as well as the effects of collaborative search tool usage on query behavior for users involved in collaborative web searches. In addition, this study aims to identify the influence of the CIS process on collaborators' querying strategies (Al-Sammarraie & Umapathy, 2016). Controlled experiments have been conducted to empirically investigate research questions and the imposed hypothesis. The experiments are conducted on groups of users that were randomly assigned to either the experimental group or the control group. The participants were divided into teams of two. Each team was randomly assigned to use either a collaborative search tool (SearchTeam), or a traditional search engine (Google).

Test Condition	Independent Variable	Control Variables	Control Procedure	Dependent Variables
Experimental	Collaborative Search Tool (SearchTeam)	Participant's level of expertise: Novice or Expert.	• Random assignment of subjects to use collaborative search tool (SearchTeam) or traditional search engine (Google). • Random pairing of novice subjects with an expert or another novice.	• Query Success Rate. • Task Completion Time. • Query Sources of Terms. • Query Patterns.
Control	Google Integrated Platform			

Table 3: Experimental Variables

3.3 SearchTeam: a Collaborative Search System

SearchTeam (http://searchteam.com/) is a free web search system that can be used by a single user or a group of users. Similar to other collaborative search systems, it supports explicit collaborative search. The system offers both search and collaboration features. The system displays list of teammates, and provides a chatting space for team users. This chat box will be hidden for individual users (Al-Sammarraie & Umapathy, 2016).
The main page has three different tabs that appear in the right top corner (see Figure 1). The *Home* tab is used to display the search screen for a single user. The *My SearchSpaces* tab is used to display the search screen where users can save search results, invite teammates and chat with team members. The *Create New* tab can be used to create either a new collaborative workspace or a personal research, or just perform a quick search. Participants are expected to use the second and third tabs to work together on search tasks. The main interface of My SearchSpaces contains three areas: the topic of the search space, the search area, and team management area (Al-Sammarraie & Umapathy, 2016). The search area and the team management area in SearchSpaces are used to support aspects of collaborative search.

Among the features that have been presented in section 2.1.2, SearchTeams supports: awareness, knowledge sharing, and the division of labor. The search area supports the awareness aspect by allowing users to view each other's saved web sites, posted comments, and past searches. Users can also use the search area to conduct the web search process. SearchTeam uses Google Search engine behind the scenes to retrieve the

search results. The manage team area supports both the knowledge sharing as well as the division of labor aspects by facilitating communication among collaborators using the team chat feature.

SearchTeam can be classified along four dimensions that we presented in section 2.1.3. First, depth of mediation, SearchTeam acts as passive element helping with aspects such as communication and awareness only. Second, intent, since SearchTeam supports explicit collaboration, therefore it is considered as a collaborative system. Third, concurrency, SearchTeam can support both synchronous and asynchronous collaborative activities. Lastly, location, SearchTeam can support both co-located and remote collaborators.

Very few of the systems that have been presented in the literature are available for outside use. For example, SearchTogether offers different features supporting CIS aspects such as awareness and division of labor; however, this tool is unavailable commercially, and is mainly used with the research community (Kelly & Payne, 2014) . Additionally, the other freely available collaborative search tools, such as Coagmento (http://coagmento.org/) do not offer the web searching process as part of their interface. This tool requires collaborators to install a plugin to their browsers prior to start the collaborative search process.

All these reasons made SearchTeam an appropriate choice for our study. SearchTeam is a freely available search engine that offers a variety of features designed to support

aspects of CIS (awareness, division of labor, and knowledge sharing) within one interface without the need to install any browser plugins.

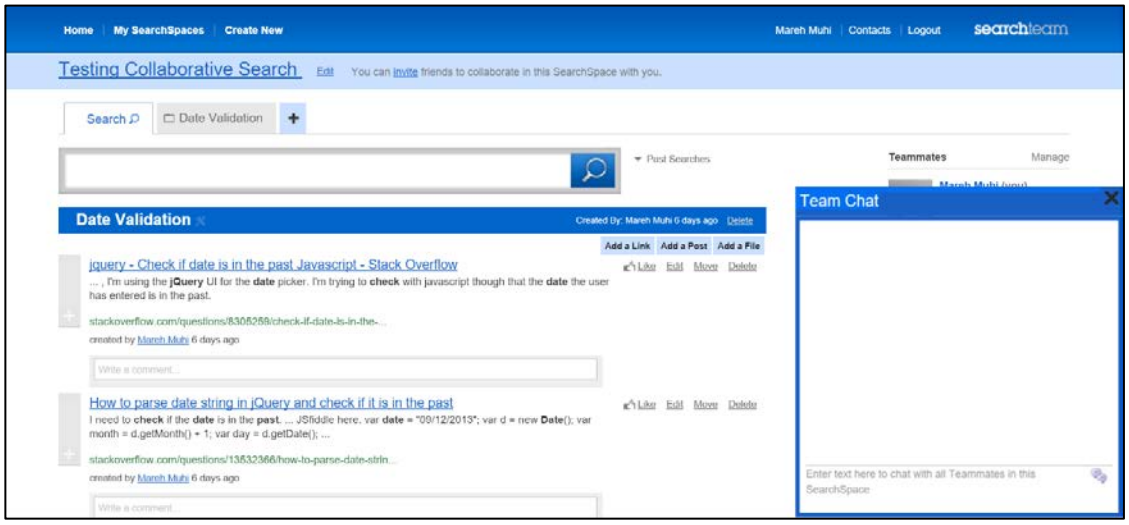


Figure 1: My SearchSpaces with Team Chat View

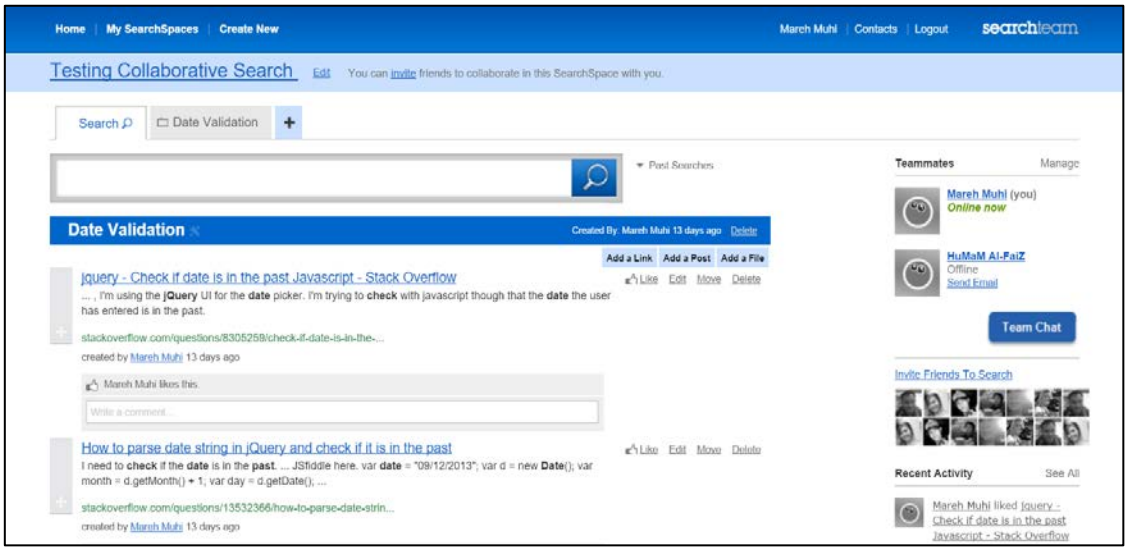


Figure 2: Shared Workspace for Saved Search Results

3.4 Google Integrated Platform

Google integrated platform is a suite of cloud-based services. It includes file storage and sharing with Google Drive, real-time collaboration with Google Docs, video meeting with Google Hangouts, and email with Gmail. Google applications can be combined to support general aspects of the CIS process such as awareness, division of labor, and knowledge sharing.

In this study, Google Drive and Google Docs are used to facilitate the real-time collaboration and sharing of search results among collaborators. This shared workspace is required to support the aspect of awareness in CIS process. Google Hangouts is used to help collaborators establish division of work and to enable sharing of knowledge through communication.

3.5 Experimental Conditions

This thesis used 2×2 between-subject design (Shadish, Cook, & Campbell, 2002) to study the differences in collaborators' search behavior when they are using a collaborative web search tool (SearchTeam) versus a traditional search engine (Google), and the impact of the collaborative information seeking process between domain experts and novices on query reformulations (Al-Sammarraie & Umapathy, 2016). In this experimental design, novice-expert teams and SearchTeam CIS tool are the treatment groups. In order to study the treatment effects, control groups are needed to act as a baseline comparison. Novice

with another novice are paired for baseline comparison to novice-expert pairs. Integrated Google platform is used for baseline comparison of SearchTeam CIS tool. Collaborative team type conditions (see Table 4) are organized into four conditions and as follows:

1) Group A – SearchTeam and Expert-Novice Collaborative Team condition: In this condition, two participants (one is an expert in the domain knowledge and the other is a non-expert) formed a team that worked on the same task at the same time. This group teams have used a collaborative web search tool (SearchTeam) to finish their assigned tasks. To simulate remotely-located collaboration, team members were told to only communicate with each other by using chat or reading each other's search histories and saved information. The collected results were shared and available in the team's workspace.
2) Group B – Traditional and Expert-Novice Collaborative Team condition: Similar to Group A, this condition also has two participants – one expert and one non-expert – and they worked together on the same task at the same time. However, Group B teams have used a traditional search tool (Google) to complete their assigned tasks. To simulate remotely-located collaboration, team members were told to only communicate with each other using Google Hangout or reading each other's search histories and saved information in Google Docs.
3) Group C – SearchTeam and Novice-Novice Collaborative Team condition: Similar to Group A, participants have used SearchTeam to complete their tasks. However, in this condition, participants were a pair of novices, and they worked together on the same task at the same time.

4) Group D – Traditional and Novice-Novice Collaborative Team condition: Similar to Group B, in this condition participants have used a traditional search tool (Google) to complete their assigned tasks. Similar to Group C, participants in this condition were a pair of novices.

Collaborative Teams	**Collaborative Search Systems**	
	SearchTeam	***Google***
Novice-Expert Team	Group A	Group B
Novice-Novice Team	Group C	Group D

Table 4: Experimental Conditions

A baseline comparison to investigate query performance improvements of a novice when paired an expert and both uses a CIS tool is needed. Novice with another novice are paired for baseline comparison purpose.

3.6 Search Tasks

Many empirical studies used two different task types to show the impact on query reformulation patterns: objective (recall-oriented and information-gathering task) and subjective (utility-based and decision-making task) (Yue et al., 2013). These two tasks represent two different types of exploratory web search tasks (Yue et al., 2013). According to the same author, the relevance criteria for selecting a search result in these two tasks are not similar. In recall-oriented and information-gathering tasks, the topical relevance is the most important criterion for selecting a search result, because the whole task is objective in relevance judgments. Whereas, in utility-based and decision-making

tasks, the relevance criteria are subjective, and they involve users' subjective judgment and personal preference.

The objectives of this research are to study query reformulation patterns as well as collaborative search tool usage on query strategies. Thus, similar to the past study (Monchaux et al., 2015), the focus was only on objective task complexity to investigate the effects of novice-expert paired teams on query strategies required to solve problems (Al-Sammarraie & Umapathy, 2016).

3.7 Experimental Search Task Context

In this experimental search task scenario, an intern (novice) is working with a mentor (expert) to solve technical software development problems. Novice-expert paired teams, as well as novice-novice paired teams have solved ASP.NET MVC problems (see Appendix A for tasks descriptions that was presented to participants). The classic CIS configuration, that is built based on two factors: location and time (Shah, 2014), has been followed. Thus, the paired teams have worked together remotely, but at the same time, to identify relevant solutions for the problems. The paired teams have searched for relevant solutions using appropriate query terms and selected a result item as the most relevant solution that they found for the given problem. The teams were encouraged to review each other's query terms and engage in an iterative search process to reach consensus on which result item describes the most ideal solution. While participants are provided with source code, they were not expected to implement the code and solve the problems.

In particular, a Bookstore Management App is used as an example for collaboratively seeking solutions for the search tasks. Bookstore Management App is a web application that automates all operations of an online book store. Using this application, users could search the online book catalog and place an order for a book. The store manager could manage the book catalog, and review and fulfill book orders (Al-Sammarraie & Umapathy, 2016).

3.8 Experiment Procedure

Because this study utilized human subjects as a part of the investigation, the project was submitted to the UNF Institutional Review Board (IRB). The study presented no risk to participants and therefore qualified for expedited IRB review. The study was approved on November 1st, 2016. The IRB reference number for this project is 962132-2 (see Appendix D).

During the experiment, students, as well as senior programmers who were willing to participate voluntarily, were first given general information about the search tasks, and were asked to read and sign the consent form for participating in the experiment. Participants were randomly divided into teams. Then teams were randomly assigned to work with Google or SearchTeam. Participants were required take part in the study at the UNF campus. Team members were placed in different rooms to simulate a remote-synchronous collaborative condition. The researchers have briefly explained the tasks and demonstrated how to use the search tools and how to save search results. Participants

were also given a list of instructions on how to use the search tools and how to save search results (see Appendix B). Participants were provided enough time to familiarize themselves with their tasks and the interfaces before they started the experiment. During this time the experimenter was available to answer any questions they may have had. Participants have worked together on their tasks and were given a maximum of 1.5 hours to finish all tasks.

The Snagit recorder is used to record all on-screen actions of the participants, including information searching, bookmarks, and URL clicks. After finishing the search task, the participants were presented with a post-search questionnaire (see Appendix C for the full survey). No personal or identifiable data on the questionnaire are collected. Participants were allowed to skip a task if that task proved too difficult to resolve.

3.9 Study Participants

The participant population is consisted of two groups: novices and experts. The first group, novice participants, included UNF School of Computing students who were over the age of 18 and had less than a year of work experience in the software development field. The second group, experts, included senior software developers, from local companies, who were over the age of 18 and had 5 to 15 years of work experience in the software development field, specifically with ASP.NET. Participants are requested to take part in the study voluntarily. A help request is sent to School of Computing professors at UNF, who have taught junior and senior level courses, to post information

about the study in their classes and request their students to voluntarily take part in the study. The researchers have relied on their network to find and contact senior programmers from local companies, and have also used LinkedIn to post information about the study.

Chapter 4

DATA ANALYSIS METHODS AND EXPECTED RESULTS

This study focused on analyzing query behaviors of users during the process of collaborative information seeking. Data were collected from transaction logs such as users' queries (search terms), search results selected as a solution, users' actions, such as procedures followed to come up with relevant results collaboratively, and text messages exchanged via chat tools. The data gathered were analyzed in two dimensions. First, quantitatively which includes two measurements: query success rate and task completion time. Second, qualitatively and in which query sources of terms, query reformulation patterns, and survey results are analyzed.

4.1 Quantitative Measures

The following measurements are calculated:

1) Query success rate (QSR) measurement: The same criterion presented in (Yue et al., 2013), to calculate query performance rate, is followed. The query is considered successful if search result items are saved after a query is issued, after subsequent discussions between team members, and the saved results are relevant to the correct answers. Using log data for each collaborative team, the number of successful queries for each collaborative team is calculated, then the average success rates is compared

for each group in each task. QSR is calculated as a ratio of number of queries with items saved and overall number of queries issued by the team.

$$\text{QSR} = \frac{\text{Queries with items saved}}{\text{Total no.of queries}}$$

Equation 1: Query Success Rate (QSR)

Our hypothesis for this metric was that collaborative teams who used tools designed for the CIS process have a higher QSR compared to other collaborative teams who used Google integrated platforms.

2) Task completion time (TCT) (in minutes): For each task, the time it takes each team to perform a search session and save relevant search results is measured. This measurement has only been calculated for teams who have accomplished one or more tasks successfully. Our hypothesis for this metric was that collaborative teams that used tools designed for the CIS process require less time to search for both simple and complex tasks with higher query success rates. Task completion time for a task is calculated as the ratio of time spent by each team to finish a task and overall number of teams who accomplished the task.

$$\text{Total Average TCT} = \frac{Time\ spent\ by\ team1 + Time\ spent\ by\ team2 + \cdots Time\ spent\ by\ team\ N}{Total\ no.of\ teams\ (N)}$$

Equation 2: Task Completion Time (TCT)

4.2 Qualitative Measures

1) Query sources of terms: The screen records and chat logs for the four team conditions were reviewed to identify possible sources of successful query terms. As part of the analysis, the successful queries for novice participants were considered and the remaining issued queries were excluded. To identify the possible query sources, first we collected all the successful queries issued by novices in each team. Then, we looked into the keywords in each query and identified the source of each of these terms. At the end, queries are classified based on the identified sources. Our hypothesis for this metric was that novices in the four collaborative team conditions might have two types of queries: self-issued and queries issued with help from an expert or novice partner. However, the number of queries issued with help from a partner is higher in expert-novice pairs who used CIS tool.
2) Query reformulation patterns: The classification of query reformulation presented in (Yue et al., 2013) is used, see (Table 5). Two consecutive queries (Q1, Q1+1) issued in the same search session will be considered as a query reformulation pattern. Each pattern issued by novice members is classified and tallied accordingly. Query reformulations are analyzed to study differences in patterns issued by novices. Patterns are analyzed to study the impact of collaborative search tools on improvements in query reformulation strategies among novices over time (Al-Sammarraie & Umapathy, 2016).

Type	Definition
New (N)	If Q1 is the first issued query, and does not share any common terms with Q1+1
Generalization(G)	Q1 and Q1+1 share common terms, and Q1+1 contains fewer terms than Q1
Specialization (S)	Q1 and Q1+1 share common terms, and Q1+1 contains more terms than Q1
Reconstruction (R)	Q1 and Q1+1 share common terms, and Q1+1 has the same length as Q1

Table 5: Query Reformulation Types

Our hypothesis for this metric was that novices in collaborative teams who used tools designed for the CIS process are more likely to use new and specialized terms more often in their queries.

Chapter 5

EVALUATION RESULTS AND ANALYSIS

For this study, the following qualitative and quantitative measures were gathered by screen recording participants' actions: query success rate, sources of query terms, query reformulation patterns, and task completion time. Data were recorded for each collaborative team condition. Using log data for the 13 teams, researchers calculated the number of successful queries for each collaborative team, then they compared the average success rates for each group in each task. Detailed analysis for each group condition is provided in section 5.1. Then researchers looked at the sources of terms for novices' successful queries. They classified novices' successful queries into two types: self-issued queries and queries issued with help from an expert partner. Subsequently, they calculated the total number of each type and compared the results for each group. The analysis results for the sources of terms is provided in section 5.2. In addition to the success rate and the sources of term measures, the researchers have studied the query reformulation patterns for novices' successful queries. The results of each reformulation type for each group condition are provided in section 5.3. The researchers have also calculated total time each collaborative team took to complete all four tasks and compared the time for each group to determine the amount of overhead tied to the use of each tool. The results of task completion time are provided in section 5.4. Additionally, the survey responses and results are provided in section 5.5. Finally, final discussions on the experiment results are provided in section 5.6.

5.1 Query Success Rate

For Group A condition, the total average of successful query rate is higher for all four tasks compared to other conditions presented below. A breakdown of the successful query rates for Group A is provided in Table 6.

	Teams	Total Number of Successful Queries Issued by a Team	Total Number of Queries Issued by a Team	Total Average Successful Query for Each Team/Task	Total Average Successf-ul Query for Each Task
Task1	Expert – Novice-A-t1	3	4	0.75	0.5
	Expert – Novice-A-t2	3	7	0.42	
	Expert – Novice-A-t3	4	12	0.33	
	Expert – Novice-A-t4	4	8	0.5	
Task2	Expert – Novice-A-t1	3	6	0.5	0.48
	Expert – Novice-A-t2	4	11	0.36	
	Expert – Novice-A-t3	2	3	0.66	
	Expert – Novice-A-t4	3	7	0.42	
Task3	Expert – Novice-A-t1	1	1	1	0.27
	Expert – Novice-A-t2	1	9	0.11	
	Expert – Novice-A-t3	0	6	0	
	Expert – Novice-A-t4	0	3	0	
Task4	Expert – Novice-A-t1	1	7	0.14	0.30
	Expert – Novice-A-t2	1	4	0.25	
	Expert – Novice-A-t3	1	2	0.5	
	Expert – Novice-A-t4	2	6	0.33	

Table 6: Query Success Rate for Group A

For the first and second tasks, the total average of the successful query rate for all four teams is 50% and 48%, which indicates that about half of the queries issued by each team were followed by at least one relevant saved item. For the third and fourth tasks, the total average of successful query rate for all four teams is 0.27% and 0.30%, which indicates that about one-third of the queries issued by each team were followed by at least one

relevant saved item. The high rate of successful queries may be due to the fact that expert individuals are part of the collaborative team, and they may have either supported novices' search efforts by providing some sort of clarification about the tasks, or performed the searching and saving of accurate results themselves.

For Group B condition, the total average of successful query rate is high for three tasks only. The third task has the lowest success rate. Breakdown of the successful query rates is provided in Table 7.

	Teams	Total Number of Successful Queries Issued by a Team	Total Number of Queries Issued by a Team	Total Average Successful Query for Each Team/Task
Task1	Novice – Novice-B-t1	4	7	0.5
Task2	Novice – Novice-B-t1	4	7	0.5
Task3	Novice – Novice-B-t1	0	0	0
Task4	Novice – Novice-B-t1	1	3	0.3

Table 7: Query Success Rate for Group B

Group B condition has only one team (Expert - Novice t1). This team found relevant answers to three tasks out of four. For the first and second tasks, the total average of successful query rate is 50%, which indicates that about half of the queries issued by each team were followed by at least one relevant saved item. For third task, the total average successful query rate is 0, which indicates that the team could not find relevant answers, or simply has skipped the task. For the fourth task, the total average successful query rate is 30%, which indicates that about one-third of the queries issued by each team were followed by at least one relevant saved item.

For Group C condition, the number of successful queries is low for all four tasks. None of the teams found relevant answers for all tasks. Only one team (Novice - Novice t1) out of three teams achieved the first task and the first part of the second task. For the first task, the total average of successful query rate is 22%, which indicates that about one-fifth of the queries issued by each team were followed by at least one relevant saved item. For the second task, the total average successful query rate is 10% only (see Table 8).

The low successful query rate may be caused by the fact that participants in Group C have limited knowledge about the search topic, and participating in a collaborative search with a novice partner using a collaborative Web search tool did not impact their query performance.

	Teams	Total Number of Successful Queries Issued by a Team	Total Number of Queries Issued by a Team	Total Average Successful Query for Each Team/Task
Task1	Novice – Novice-C -t1	2	9	0.22
	Novice – Novice-C-t2	0	9	0
	Novice – Novice-C- t3	0	6	0
Task2	Novice – Novice-C- t1	1	10	0.1
	Novice – Novice-C- t2	0	11	0
	Novice – Novice-C- t3	0	5	0
Task3	Novice – Novice-C- t1	0	3	0
	Novice – Novice-C- t2	0	9	0
	Novice – Novice-C- t3	0	5	0
Task4	Novice – Novice-C- t1	0	3	0
	Novice – Novice-C- t2	0	8	0
	Novice - Novice -C-t3	0	8	0

Table 8: Query Success Rate for Group C

For Group D condition, the number of successful queries is low for all four tasks. None of the five teams found relevant answers for all tasks. Only two teams (Novice - Novice

t1 and Novice - Novice t5) out of five achieved part of the first and the second tasks. A breakdown of the successful query rates for Group D teams is provided in Table 9.

	Teams	Total Number of Successful Queries Issued by a Team	Total Number of Queries Issued by a Team	Total Average Successful Query for Each Team/Task
Task1	Novice – Novice-D-t1	1	4	0.25
	Novice – Novice-D-t2	0	2	0
	Novice – Novice-D-t3	0	4	0
	Novice – Novice-D-t4	0	5	0
	Novice – Novice-D-t5	0	3	0
Task2	Novice – Novice-D-t1	1	3	0.33
	Novice – Novice-D-t2	0	2	0
	Novice – Novice-D-t3	0	4	0
	Novice – Novice-D-t4	0	2	0
	Novice – Novice-D-t5	1	3	0.33
Task3	Novice – Novice-D-t1	0	3	0
	Novice – Novice-D-t2	0	0	0
	Novice – Novice-D-t3	0	2	0
	Novice – Novice-D-t4	0	2	0
	Novice – Novice-D-t5	0	2	0
Task4	Novice – Novice-D-t1	0	2	0
	Novice – Novice-D-t2	0	1	0
	Novice – Novice-D-t3	0	2	0
	Novice – Novice-D-t4	0	3	0
	Novice – Novice-D-t5	0	2	0

Table 9: Query Success Rate for Group D

For the first task, the total average successful query rate is 25%, which indicates that about one-fourth of the queries issued by each team were followed by at least one relevant saved item. For the second task, the total average of successful query rate is 30%. The low successful query rate may be caused by the fact that participants in Group D have limited knowledge about the search topic, and participating in a collaborative search with a novice partner using Google Integrated platforms did not influence their query performance.

5.2 Sources of Terms (for Novices' Successful Queries)

For Group A collaborative teams, there are two types of sources that novice participants used to come up with successful queries. First, they used self-issued terms. Second, they also incorporated technical keywords provided by expert partners via a chat tool. Our analysis shows that most of the successful queries were self-issued by the novice participants (87.5%), and the remaining were from keywords provided by the experts (12.5%) (see Table 10).

Source of Query Terms	Number of Instances	%
Self-Issued	14	87.5
Issued with help from an expert	2	12.5
Total	16	100

Table 10: Sources of Query Terms for Group A

Table 11 shows that novice participants in three teams out of four have at least one query that is generated with the help of an expert partner. This indicates that the novice-expert pairs may have worked closely on the same task at the same time, and that experts had a chance to mentor and assist the novices whenever they felt it necessary.

	Teams	Total Number of Successful Queries Issued by a Student	Source of Terms (for successful queries issued by a novice)
Task1	Expert – Novice-A-t1	1	Self-issued
	Expert – Novice-A-t2	1	Self-issued
	Expert – Novice-A-t3	6	Self-issued
	Expert – Novice-A-t4	1	Self-issued
Task2	Expert – Novice-A-t1	1	Issued with help from an expert
	Expert – Novice-A-t2	2	Self-issued query and Issued with help from an expert
	Expert – Novice-A-t3	2	Self-issued query and Issued with help from an expert
	Expert – Novice-A- t4	1	Self-issued
Task3	Expert – Novice-A- t1	None	Failed to find relevant result
	Expert – Novice-A- t2	None	Failed to find relevant result
	Expert - Novice-A- t3	None	Failed to find relevant result
	Expert – Novice-A-t4	None	Failed to find relevant result
Task4	Expert – Novice-A- t1	1	Self-issued
	Expert – Novice-A- t2	None	Failed to find relevant result
	Expert – Novice-A- t3	1	Issued with help from an expert
	Expert – Novice-A- t4	1	Self-issued

Table 11: Successful Queries for Group A (Novices Participants)

Through the analysis of chat logs and query logs for Group A collaborative teams, the researchers found that all experts in four teams worked closely with their novice partners to ensure that only relevant results were considered and saved. For example, one expert advised a novice partner to focus the search on a certain topic by providing hints on what to search for. The expert asked the novice partner to look for specific topics about routing in MVC.

In the chat log:

> **Expert**: so we need routing search results, but that is a big topic, we can search for specifics inside routing.

Novice: ok let search for routing.

The word “routing” is introduced by the expert and the student borrowed it to form the following successful query:

“mvc routing based on role”

Another expert asked a novice partner to look for an online resource that might help solving the problem. In the chat log:

Expert: For something like this, I would recommend looking at MSDN or something close to the official Microsoft Docs on Routing and the RouteConfig.

Novice: never really put 'microsoft' in the search; but in doing so found that other page I saved. Good to know (for "official" answers)

The word “microsoft” is introduced by the expert and the student borrowed it to form the following successful query:

“mvc asp.net routing microsoft”

Another expert helped a novice partner with the wording of the search query by suggesting an example query.

In the chat log:

Expert: maybe look for "mvc how to authenticate"

Novice: gotcha

The novice used the provided text and put it in the following successful query:

“mvc asp.net how to authenticate”

For Group B collaborative teams, the novice participant used self-issued terms only. The novice did not use any technical keywords provided by the expert during task discussions that occurred via the chat. The analysis shows that the novice participant had a total of 5

successful queries for the first and the second tasks, and these queries were all self-issued (see Table 12). This indicates that the novice-expert pair did not work on the same task at the same time. Instead they may have followed the brute force strategy, where they searched separately and whoever come up with an interesting result sent it to the other, and results were merged afterwards. One possible explanation is that expert participant was not fully aware of the activities of the novice partner during the search task. Unlike participants who used SearchTeam, which provides alerts whenever a partner saves a result into the shared workspace. Participants who used Google docs don't have such an awareness feature, and they had to toggle between open windows for the entire search session.

	Teams	Number of Successful Queries Issued by Novices	Source of Terms (for successful queries issued by a novice)
Task1	Expert – Novice-B-t1	2	Self-issued
Task2	Expert – Novice-B-t1	3	Self-issued
Task3	Expert – Novice-B-t1	Failed to find relevant result	
Task4	Expert - Novice -B-t1	Failed to find relevant result	

Table 12: Successful Queries for Group B (Novice Participant)

For Groups C and D collaborative teams, all the terms used in the successful queries were self-issued. The source of these keywords was from the participants themselves. The participants didn't use terms exchanged via the chat tool. Instead they either used words from the proposed questions, or employed terms they had from previous knowledge. This indicates that most team members did not work on the same task at the same time. Instead they may have followed the brute force strategy, where they searched separately

and whoever come up with an interesting result sent it to the other, and results were merged afterwards.

5.3 Query Patterns (for Novices' Successful Queries)

For Group A teams, the results indicate that the novice participants in this group condition tended to use the Specialization pattern for self-issued queries, and New pattern for queries issued with an expert's assistance. One possible explanation is that novices in this group are more likely to receive advice/guidance from an expert partner about how to approach a search task. During collaborative discussions, novices may have either picked terms to make their self-issued queries more specific, or used wording suggestions offered by their expert partners to issue new queries. Table 13 shows the finding of query reformulation patterns for Group A condition.

	Teams	Total Number of Successful queries issued by a Student	Source of Terms (for successful queries issued by a student)	Query Patterns
Task1	Expert – Novice-A-t1	1	Self-issued	New
	Expert – Novice-A-t2	1	Self-issued	New
	Expert – Novice-A-t3	6	Self-issued	New + Specialization
	Expert – Novice-A-t4	1	Self-issued	New
Task2	Expert – Novice-A-t1	1	Issued with help from an expert	New
	Expert – Novice-A-t2	2	Self-issued query	Specialization
			Issued with help from an expert	New
	Expert – Novice-A-t3	2	Self-issued query	New
			Issued with help from an expert	New
	Expert – Novice-A-t4	1	Self-issued	New
Task3	Expert – Novice-A-t1	None of the novices in the four collaborative teams could find relevant answer to this task		
	Expert – Novice-A-t2			
	Expert – Novice-A-t3			
	Expert – Novice-A-t4			
Task4	Expert – Novice-A-t1	1	Self-issued	New
	Expert – Novice-A-t2	Failed to find relevant result		
	Expert – Novice-A-t3	1	Issued with help from an expert	Reconstruction
	Expert – Novice-A-t4	1	Self-issued	New

Table 13: Query Reformulation Patterns for Group A (Novices Participants)

For Group B teams, the results show that the successful queries of the novice participant were all self-issued, and the New pattern is the most used query reformulation strategy. It may be due to the fact that the novice participant had some knowledge about parts of the search topic, thus that same participant retrieved the correct results without performing many of query reformulation steps.

For Group C collaborative teams, only one team (Novice - Novice t1) out of the three teams was able to reach correct results for the first task and part of the second task. Reconstruction was the most frequently used reformulation type by the collaborative team. A possible explanation is that, since participants in this team have less knowledge about the search topic, they may have used the Reconstruction strategies frequently to explore the topic in depth.

For Group D collaborative teams, only two teams (Novice - Novice t1 and Novice - Novice t5) out of five teams were able to reach correct results for part of the first and second tasks. New was the most frequently used reformulation type by the collaborative team. This may be due to the fact that participants in these two teams had some knowledge about parts of the search topic, thus they could retrieve the correct results without performing many query reformulation steps.

5.4 Task Completion Time

The task completion time is calculated only for teams who found relevant solutions to the proposed problems. For group A teams, only two teams out of four have finished all four tasks. The teams in this group took longer to finalize the first and second tasks. On average, it took Group A collaborative teams around 27 minutes to finish the first task and 18.75 minutes to finish the second task. This may be due to the time experts took to help explain some points and validate saved results by their novice partners. Table 14 summarizes all the findings of time spent by each team to complete search tasks.

	Teams	Task Completion Time in (minutes)	Total Average for Each Task in (minutes)
Task1	Expert – Novice-A-t1	25	27 min
	Expert – Novice-A-t2	24	
	Expert – Novice-A-t3	28	
	Expert – Novice-A-t4	31	
Task2	Expert – Novice-A-t1	13	18.75 min
	Expert – Novice-A-t2	24	
	Expert – Novice-A-t3	13	
	Expert – Novice-A-t4	25	
Task3	Expert – Novice-A-t1	4	8 min
	Expert – Novice-A-t2	12	
	Expert – Novice-A-t3	Failed to find relevant result	
	Expert – Novice-A-t4	Failed to find relevant result	
Task4	Expert – Novice-A-t1	12	15.75 min
	Expert – Novice-A-t2	12	
	Expert – Novice-A-t3	19	
	Expert – Novice-A-t4	20	

Table 14: Task completion Time (Group A)

Group B team had finished three tasks out of four. The team in this group took longer to finalize the three tasks. The team took around 50 minutes to finish the first task, 26 minutes to finish the second task, and 26 minutes to finalize the fourth task (see Table 15). This may be due to the time the expert participant took to work with the novice partner. In addition, we believe that since the participants in this team used Google Integrated Platforms, they needed extra time to toggle between open windows to use the chat, perform the search, or save their search results.

	Teams	Task Completion Time in (minutes)
Task1	Expert – Novice-B-t1	50 min
Task2	Expert – Novice-B-t1	26 min
Task3	Expert – Novice-B-t1	Skipped or failed to find relevant result
Task4	Expert – Novice-B-t1	26 min

Table 15: Task Completion Time (Group B)

For Group C collaborative teams, only one team (Novice - Novice t1) finished the first task and part of the second task. It took this team about 18 minutes to finalize the first task and 10 minutes to finish the second task.

For group D collaborative teams, only two teams (Novice - Novice t1 and Novice - Novice t5) out of five teams were able to find relevant results for part of the first and second tasks. The teams in this group took longer to complete the two tasks. It took the team (Novice - Novice t1) about 30 minutes to finalize the first task. For the second task, the two teams took on average of 11 minutes to complete it (see Table 16). One possible explanation is that since the participants in these two teams used Google Integrated Platforms, they needed extra time to toggle between open windows to use the chat to communicate with a partner, perform the search, or save their search results.

	Teams	Task Completion Time in (minutes)	Total Average for Each Task in (minutes)
Task1	Novice – Novice-D-t1	30	30 min
	Novice – Novice-D-t5	Failed to find relevant result	
Task2	Novice – Novice-D-t1	12	11 min
	Novice – Novice-D-t5	10	
Task3	Novice – Novice-D-t1	Failed to find relevant result	N/A
	Novice – Novice-D-t5		
Task4	Novice – Novice-D-t1	Failed to find relevant result	N/A
	Novice – Novice-D-t5		

Table 16: Task Completion Time (Group D)

5.5 Survey Responses

In addition to recording participants' actions as they went through the search tasks, we also asked the participants to complete a survey about their experience. We asked study participants some questions about their search experience using SearchTeam or Google. We collected answers from 22 participants, with 10 responses from Google users and 12 responses from SearchTeam users. In the following section, we present part of the survey results. For the full report see Appendix E.

5.5.1 Experience Rating

At the end of the search sessions we asked participants to answer a set of close-ended and open-ended questions, rating questions on a 1-5 scale based on how much they agreed with the presented statement (See Appendix C). These questions were the same for SearchTeam and Google users. Regarding the results of ranking the sources of search terms, the following four sources were rated highly by Google and SearchTeam users:

previously viewed webpages, previously saved webpages by yourself, previously issued queries by yourself, and discussion with your partner (see Figure 3). The results are consistent with the findings in section 5.2.

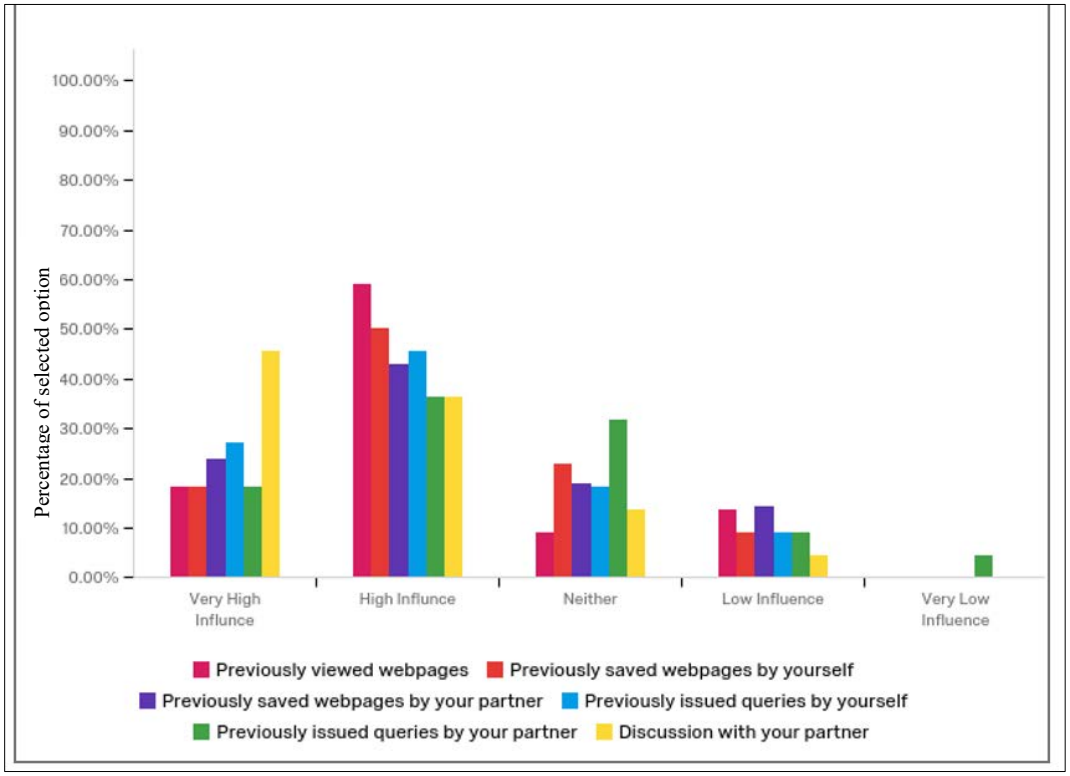


Figure 3: Survey Results - Sources of Search Terms Rating

Participants were also asked about challenges encountered related to searching. The results show that novice participants in novice-novice teams struggled with finding correct search terms and relevant answers, regardless of the search tool used. This is consistent with the findings in section 5.1. Some of the replies were:

- not knowing exactly the terms I needed to look up.
- understanding what is relevant and what isn't if you're new to programming.

- Trying to word my searches like I would if I was looking for similar solutions in the languages I commonly use.
- I was able to find answers to the given task; however I didn't always understand what I found.

One final note regarding participants' rating of the search tool. The results show that all SearchTeam users agreed that the tool brought structure and persistence to the collaboration process (See Figure 4 and 5).

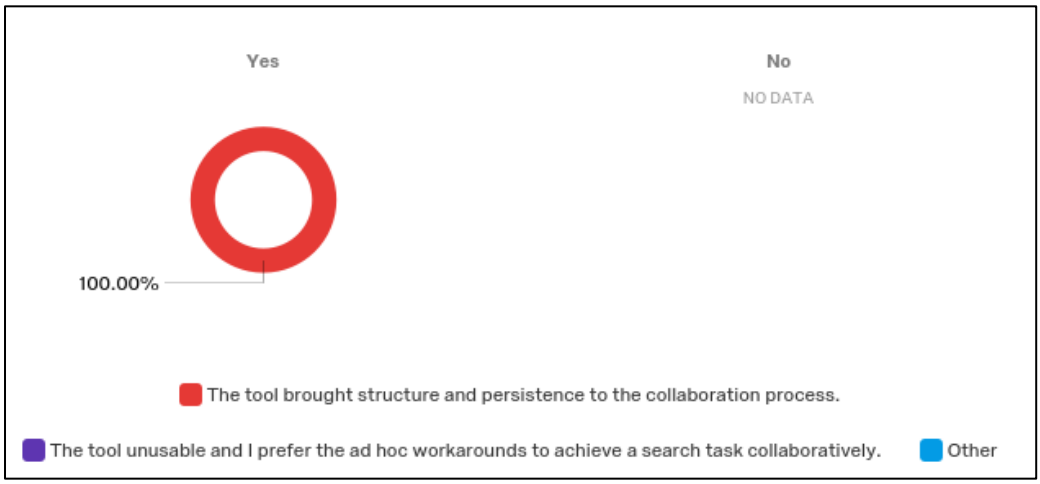


Figure 4: Survey Results - SearchTeam Rating

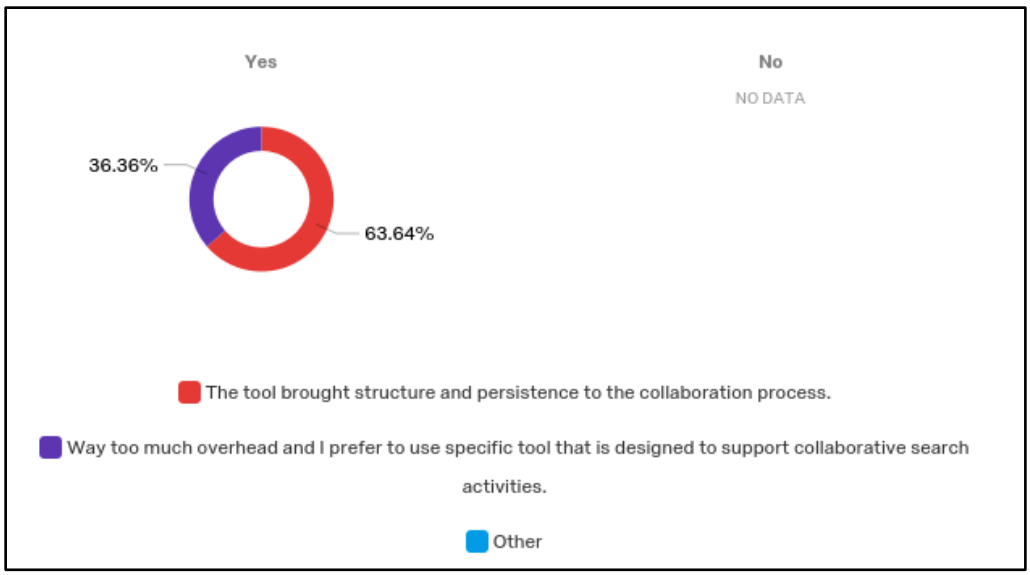


Figure 5: Survey Results - Google Platform Rating

5.6 Discussion

During the analysis of the experiment results, the researchers observed the following findings. First, by looking into the query performance for the four team conditions, they found that expert-novice pairs who used SearchTeam and Google had higher success rates compared to the novice-novice pairs. In addition, expert-novice teams who used collaborative search tools had higher success rates than expert-novice team who used Google. This suggests that the collaborative search tool may have facilitated the collaborative efforts for this type of collaborative team. However, when the researchers compared the success rates for novice-novice pairs who used SearchTeam tool with the success rates of pairs who used Google, the result wasn't the same. They noted that there isn't a big difference in the success rates for these teams. This shows that the use of collaborative search tools may not directly benefit the collaborative efforts of novice pairs who have the same level of knowledge about the search task.

Through the analysis of the sources of terms in novices' successful queries, the researchers found that novices in expert-novice team who used a collaborative search tool, included self-issued terms as well as keywords provided by their expert partners in their successful queries. In contrast, novices in expert-novice teams who used Google, only incorporated self-issued terms in their successful queries. This indicates that working with an expert using a collaborative search tool may have an impact on the way novice participants construct their successful queries. In contrast. that impact is minimal for novices who worked with experts using a traditional search engine. One possible reason is that the ease of use of the collaborative search tool, plus the awareness features provided by this tool may have allowed the experts to provide better mentoring for the novices' search activities, and offer better guidance or advice when needed. This results in more successful queries issued by novice participants. When the researchers looked at the sources of terms for novice-novice pairs who used collaborative search team and same type of pairs who used Google, they did not see any difference. Participants in both conditions either used self-issued terms or took words from the proposed questions and put them in their successful queries. This indicates that most novice participants unlikely worked on the same task at the same time, and, thus, the collaborative search activity did not impacted their way of constructing their search queries.

For the query reformulation, New and Specialization were mostly used by novices in expert-novice pairs who worked with a collaborative search tool. The researchers believe that novices may have either picked technical terms from collaborating with experts to make their self-issued queries more specific, or used wording suggestions offered by their

expert partners to issue new queries. However, the researchers found novices who worked with experts using Google platforms only used New pattern and did not incorporate technical terms provided by their expert partner. This indicates that novices preferred to perform their searches solely for familiar search topics and skip remaining tasks. A possible explanation is that the process overhead may have made it harder for collaborators to work closely and see each other search actions such as saved results, previously issued queries. Similarly, novice-novice pairs who have used Google, also used New pattern to come up with results to the search topics. This might be because, participants in these teams may have some knowledge about parts for the search topic and could retrieve the correct results without performing many query reformulation steps. In contrast, novice-novice teams who used a collaborative search tool were more likely to use Reconstruction reformulation type to explore the topic in depth, since they lacked the knowledge about the search topic.

By analyzing the task completion time, the researchers found that novice-expert teams who used the collaborative search tool required less time than novice-expert team who used Google to finalize all tasks. Similarly, it took novice-novice teams, who used the collaborative search tool, less time than what novice-novice teams, using Google required to finish all tasks. This indicates that collaborative search tools brought structure to the process and facilitated the collaborative actions. However, the researchers noted that novice-expert teams who used the collaborative search tool required more time compared to novice-novice teams, who spent less time in completing the tasks. One possible

explanation is that participants in expert-novice pairs needed to spend more time on collaboration to come up with the final results for each of the search task.

The study's researchers acknowledge some limitations. First, they have limited number of participants for each of the four conditions. The findings may be different if more people volunteered for this study. Second, they only considered novices' successful queries in their quantitative and qualitative evaluation. Analyzing all queries issued by novices, including the unsuccessful ones, may provide greater understanding about the types of actions novices perform to reach certain results with and without an expert partner. Finally, for the task completion time measurement, they only considered teams that successfully accomplished a search task or part of it.

Chapter 6

CONCLUSION

In conclusion, the purpose of this empirical study is to answer some questions regarding the effectiveness of collaborative information seeking using collaborative web search tools (Al-Sammarraie & Umapathy, 2016). The motivation for this research study is raised from researcher's personal experience as well as from interactions with local software companies. When a software developer intern is recruited by a software company, the intern is assigned to an experienced software developer. The intern is expected to learn how to troubleshoot software development problems by working with an expert. Interactions between the interns and the experts are essentially a CIS process. In this context, the question of the effectiveness of using collaborative search systems like SearchTeam, as opposed to integrated Google platform, was posed to researchers (Al-Sammarraie & Umapathy, 2016). Thus, the main research objective of the empirical study is to assess the effectiveness of the SearchTeam tool in the context of novice-expert pairs troubleshooting software development problems. The participants worked on four tasks using SearchTeam system and Google Integrated Platform under four conditions: SearchTeam and expert-novice team, SearchTeam and novice-novice team, Google and expert-novice team, and Google and novice-novice team. The results demonstrate that the successful query rate is higher in expert-novice collaborative teams who used a collaborative search tool. Participants in expert-novice collaborative teams who used a collaborative search tool, require less time to finalize all tasks compared to expert-novice collaborative teams who used a traditional search tool. Self-issued queries and chat logs

were major sources of terms utilized by novice participants in expert-novice collaborative teams who used collaborative search tool. Novices, as part of expert-novice pairs, who used a collaborative search tool, employed New and Specialization more often as query reformulation patterns. In summary, having domain experts in a collaborative team working with novices using tools designed to support CIS help increase the successful query rate, and also help novices learn and apply new terms in their successful queries.

This thesis contributes to existing literature about the possibility of employing CIS processes and tools, by providing detailed investigation regarding the influence of utilizing a collaborative search tool (SearchTeam) in the context of software troubleshooting and development. This study highlights the possible CIS activities that may occur among software developer interns and their mentors. Furthermore, this study reveals that existing features, such as awareness and built-in IM offered by SearchTeam, can promote the CIS activities among participants and help increase novices' query success rates. Finally, the study's researchers believe the use of tools designed to support collaborative search actions in software development companies, will has the potential to improve novices' query behavior and search strategies. In this thesis, the researchers provide detailed experimental design plan to study novice-expert team solving software development problems. Other researchers interested in studying software development problems in the context of CIS can use this thesis experimental design as a model to investigate their research questions.

6.1 Future Work

The researchers believe that with some additional work, this study can be extended to learn more about the query behavior of novice users. One possible way may be through the employment of larger collaborative teams in each of the team conditions presented in this study. The results of this study can also be expanded by defining and analyzing more of the quantitative measurements, such as the query vocabulary richness and query diversity for novice participants.

The study's investigators hope the research work presented in this paper inspires others to investigate collaborative web search tools and CIS process in different contexts. The results of this study show that CIS processes have a positive impact regarding augmenting querying strategies for novice users when paired with domain experts in the context of software development. Therefore, it is worth investigating the same concept, but within different domain or field setting such as chemistry, biology, etc. Lastly, this investigation may be further expanded by studying the impact of CIS actions and tools in cross-disciplinary settings.

REFERENCES

Al-Sammarraie, M., & Umapathy, K. (2016). A research plan to study impact of a collaborative web search tool on novice's query behavior. *2016 International Conference on Collaboration Technologies and Systems (CTS),* 539-544. doi:10.1109/CTS.2016.0100

Foley, C., & Smeaton, A. F. (2010). Division of labour and sharing of knowledge for synchronous collaborative information retrieval. *Information Processing and Management, 46*, 762-772. doi:10.1016/j.ipm.2009.10.010

Golovchinsky, G., Qvarfordt, P., & Pickens, J. (2009). Collaborative information seeking. *Computer, 42*(3), 47-51. doi:10.1109/MC.2009.73

Jansen, B. J., Booth, D. L., & Spink, A. (2009). Patterns of query reformulation during web searching. *Journal of the American Society for Information Science & Technology, 60*(7), 1358-1371. doi:10.1002/asi.21071

Kang, R., Fu, W., & Kannampallil, T. G. (2010). Exploiting knowledge-in-the-head and knowledge-in-the-social-web: Effects of domain expertise on exploratory search in individual and social search environments. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* Atlanta, Georgia, USA. 393-402. doi:10.1145/1753326.1753386

Kelly, R., & Payne, S. J. (2014). Collaborative web search in context: A study of tool use in everyday tasks. *Proceedings of the 17th ACM Conference on Computer Supported*

Cooperative Work & Social Computing, Baltimore, Maryland, USA. 807-819. doi:10.1145/2531602.2531617

Lewandowski, D. (2012). *Web search engine research*. Bingley, UK: Emerald Group Publishing Limited.

Mohammad Arif, A. S., Du, J. T., & Lee, I. (2014). Examining collaborative query reformulation: A case of travel information searching. *Proceedings of the 37th International ACM SIGIR Conference on Research & Development in Information Retrieval,* Gold Coast, Queensland, Australia. 875-878. doi:10.1145/2600428.2609463

Monchaux, S., Amadieu, F., Chevalier, A., & Mariné, C. (2015). Query strategies during information searching: Effects of prior domain knowledge and complexity of the information problems to be solved. *Information Processing and Management, 51*, 557-569. doi:10.1016/j.ipm.2015.05.004

Morris, M. R., Lombardo, J., & Wigdor, D. (2010). WeSearch: Supporting collaborative search and sensemaking on a tabletop display. *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work,* Savannah, Georgia, USA. 401-410. doi:10.1145/1718918.1718987

Morris, M. R., & Teevan, J. (2010). *Collaborative web search : Who, what, where, when, and why*. San Rafael, CA.: Morgan & Claypool Publishers.

Morris, M. R., Teevan, J., & Bush, S. (2008). Enhancing collaborative web search with personalization: Groupization, smart splitting, and group hit-highlighting. *Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work,* San Diego, CA, USA. 481-484. doi:10.1145/1460563.1460640

Paul, S. A., & Morris, M. R. (2009). CoSense: Enhancing sensemaking for collaborative web search. *Chi -Conference-, 3*, 1771-1780.

Paul, S. A., & Reddy, M. C. (2010). Understanding together: Sensemaking in collaborative information seeking. *Proceedings of the 2010 ACM Conference on Computer Supported Cooperative Work,* Savannah, Georgia, USA. 321-330. doi:10.1145/1718918.1718976

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference* Boston : Houghton Mifflin, 2001.

Shah, C. (2010). Collaborative information seeking: A literature review. *Advances in Librarianship, 32*(2010), 3-33.

Shah, C., Capra, R., & Hansen, P. (2014). Collaborative information seeking [guest editors' introduction]. *Computer, 47*(3), 22-25. doi:10.1109/MC.2014.54

Shah, C. (2014). Collaborative information seeking. *Journal of the Association for Information Science & Technology, 65*(2), 215.

Soulier, L., Tamine, L., & Bahsoun, W. (2014). On domain expertise-based roles in collaborative information retrieval. *Information Processing and Management, 50*, 752-774. doi:10.1016/j.ipm.2014.04.002

Stange, D., & Nurnberger, A. (2015). When experts collaborate: Sharing search and domain expertise within an organization. *Proceedings of the 15th International Conference on Knowledge Technologies and Data-Driven Business,* Graz, Austria. 45:1-45:4. doi:10.1145/2809563.2809582

Yue, Z., Han, S., He, D., & Jiang, J. (2014). Influences on query reformulation in collaborative web search. *Computer, 47*(3), 46-53. doi:10.1109/MC.2014.62

Yue, Z., Han, S., & He, D. (2013). An investigation of the query behavior in task-based collaborative exploratory web search. *Proceedings of the American Society for Information Science and Technology, 50*(1), 1-10. doi:10.1002/meet.14505001091

APPENDIX A - SEARCH TASKS

Task 1

1. The following figure is showing error messages that are automatically displayed in a view in response to invalid entries. Assume you needed to have the same kind of behavior that is triggered in response to invalid entries on one of your project's views templates. Also, you needed to ensure having this logic established for both server-side and client-side.

   Use the search tool provided to search for a solution. The source code provided can be used to help verify your search results. Follow the steps listed in the instruction document to save your search results. Please work with your partner to search for a resolution to all tasks' scenarios.

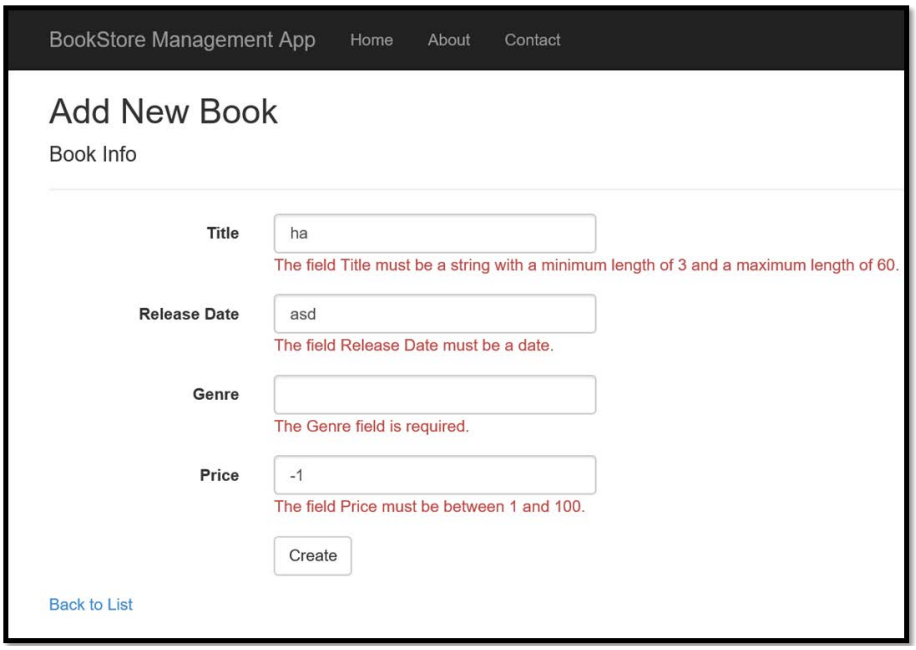


Figure 6: Add New Book View (Showing Required Validations)

2. Assume that two more additional fields (“New Release: Coming Soon”) and (“New Release: Last 30 Days”) have been added to the view as shown below. Assume you needed to have the following scenarios implemented for these two new fields:

   a. Ensure users must only enter future date in (“New Release: Coming Soon”). You would like to easily re-use the same logic across multiple screens. You need to have the logic for this scenario is implemented in server side.

   b. Ensure users must only enter dates within the past 30 days in (New Release: Last 30 Days). You would like to have this logic implemented in this screen only (reuse is not an option). You need to have the logic for this scenario is implemented in server side.

   Use the search tool provided to search for a solution. The source code provided can be used to help verify your search results. Follow the steps listed in the instruction document to save your search results. Please work with your partner to search for a resolution to all tasks’ scenarios.

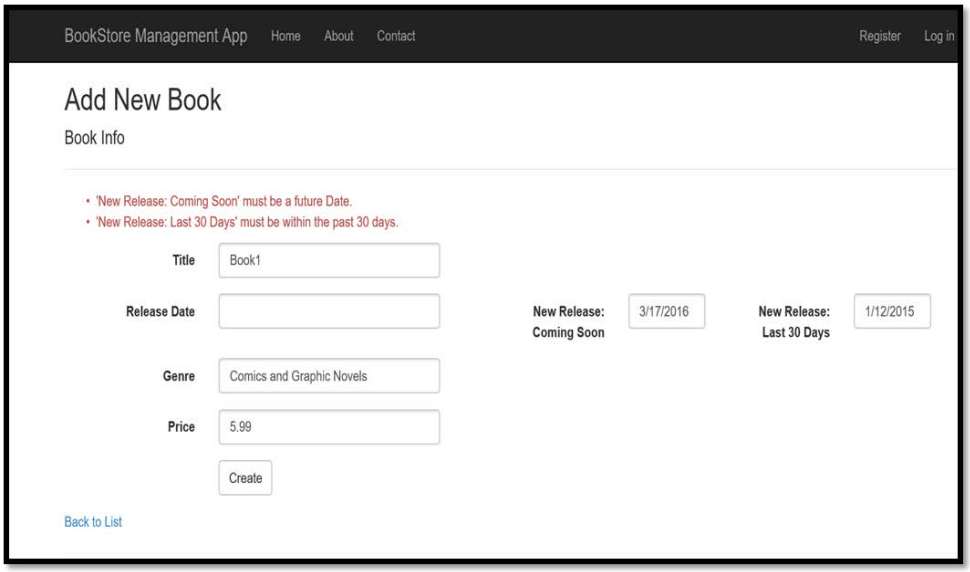


Figure 7: Add New Book View (Showing Additional Custom Validations)

Task 2

1. Suppose you have an action, called *Index ()* in the *Books* controller, that lists all saved records as shown in below figure.

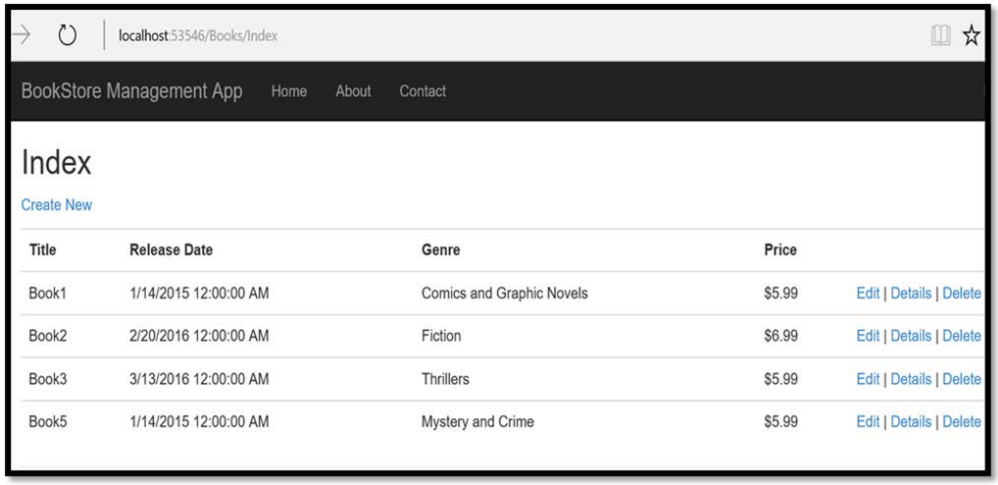
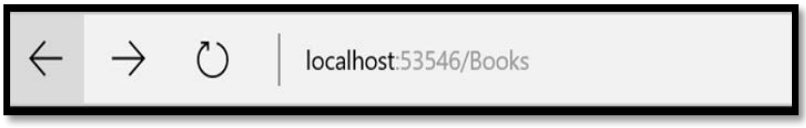


Figure 8: Index View

The URL used to direct the Get call to the *Index ()* action is defaulted as the following: http://localhost:xxxxx/Books/Index.

Assume that you've been asked to have 3 additional URLs (listed in the table below) that should be used to request the same page.

  	All three URLs should be used along with the default one to trigger the Get call and return the same view.

2. This default URL http://localhost:xxxxx/Books/Index is used to access the Index page shown in the previous task. Assume that you wanted to prevent a group of users (i.e. unauthenticated users), from accessing that page.

   Use the search tool provided to search for a solution. The source code provided can be used to help verify your search results. Follow the steps listed in the instruction document to save your search results. Please work with your partner to search for a resolution to all tasks' scenarios.

Task 3

Suppose that you got a request from your client to transform the *Genre* field into a dropdown field in Create and Edit views (see below). What would be the best approach to implement this without changing the code for that field in each and every view?

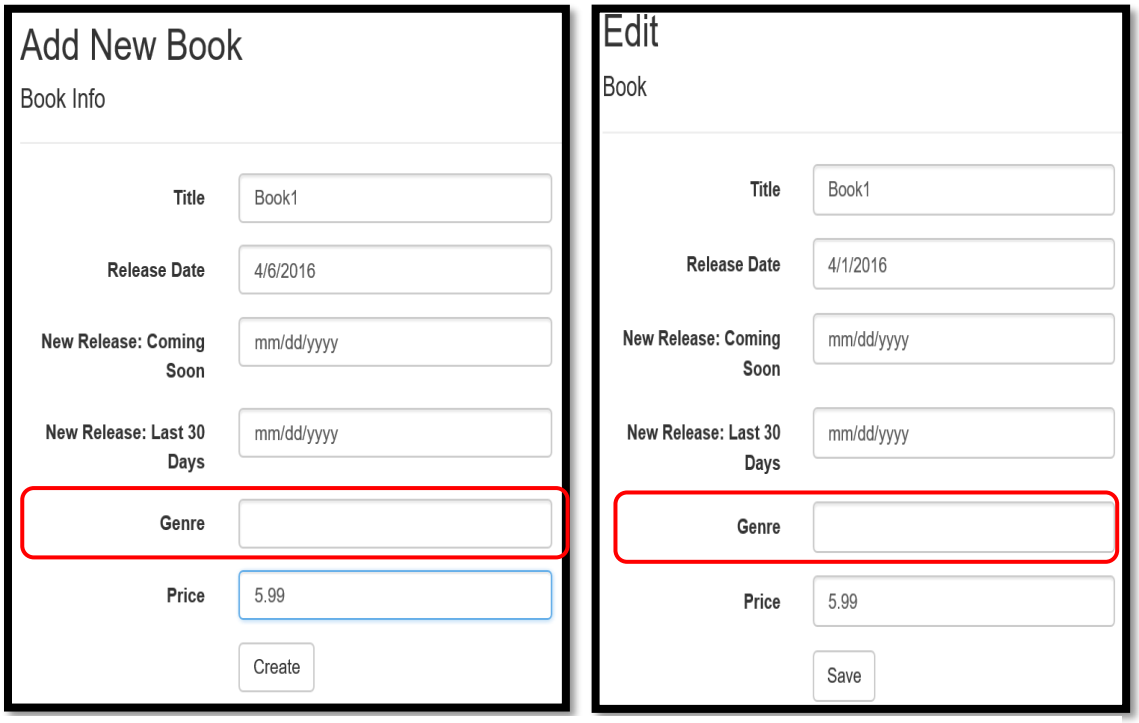


Figure 9: Add and Edit Views (Showing Genre Field Before Transformation into Dropdown)

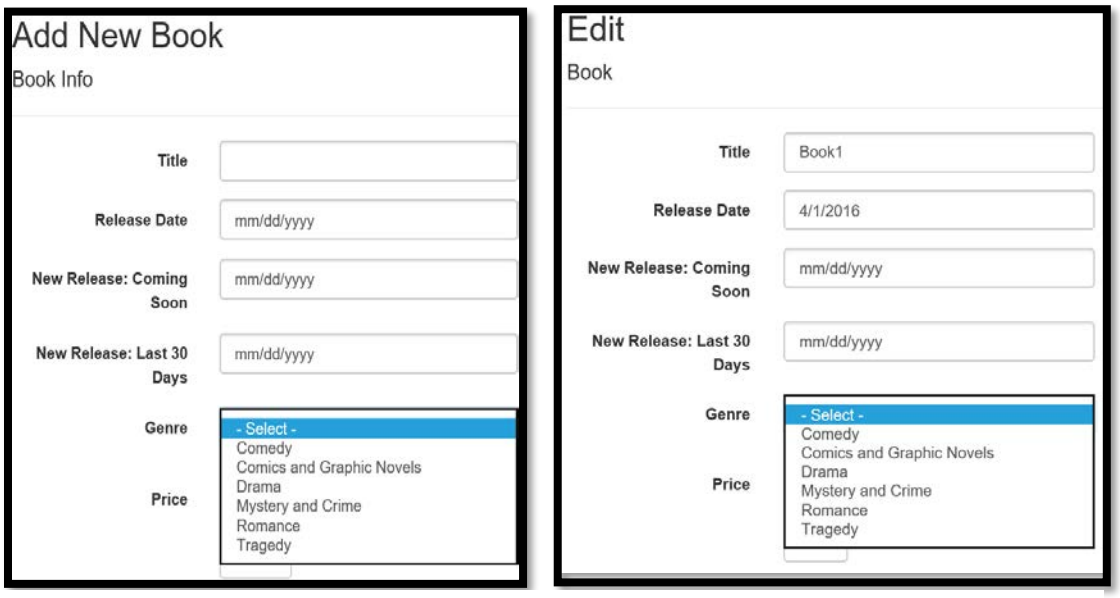


Figure 10: Add and Edit Views (Showing Genre Field After Transformation into Dropdown)

Use the search tool provided to search for a solution. The source code provided can be used to help verify your search results. Follow the steps listed in the instruction document to save your search results. Please work with your partner to search for a resolution to all tasks' scenarios.

Task 4

Assume you got a request to limit the user entries of the date type values to one when adding new book or editing the information of an existing one.
For example, if the user inserted a date in the New Release: Last 30 Days field, the two other date fields must be grayed out (see screenshots below). What would be the best approach to have this behavior in place for both views (Create & Edit) without the need to repeat the same code in each view?

Use the search tool provided to search for a solution. The source code provided can be used to help verify your search results. Follow the steps listed in the instruction document

to save your search results. Please work with your partner to search for a resolution to all tasks' scenarios.

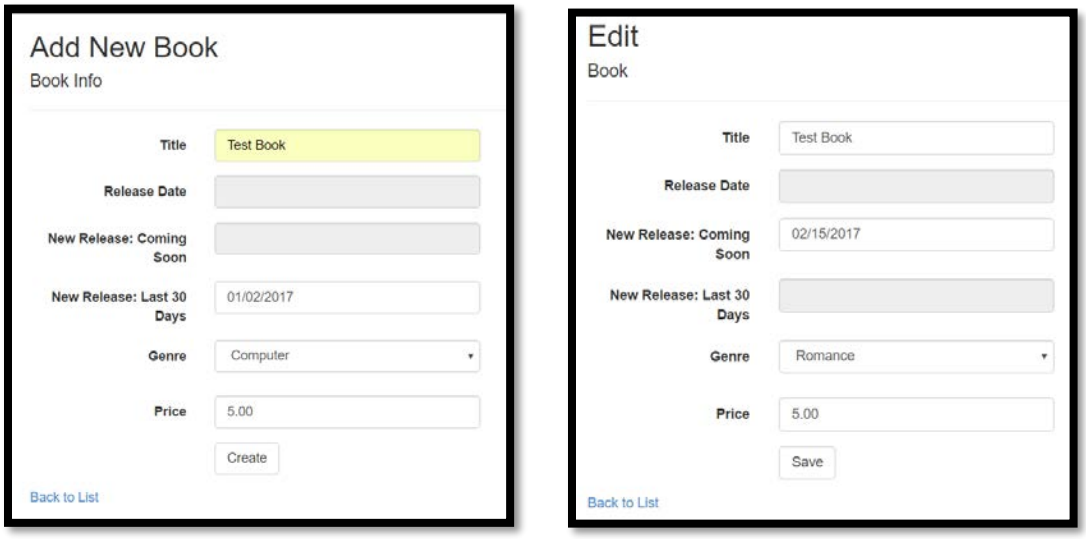


Figure 11: Add and Edit Views (Blocking the Insertion of Multiple Date Values)

APPENDIX B - LIST OF INSTRUCTIONS

Saving Search Results Instructions for SearchTeam Users

1. Click on **Task 1** (if you are working on Task 1).

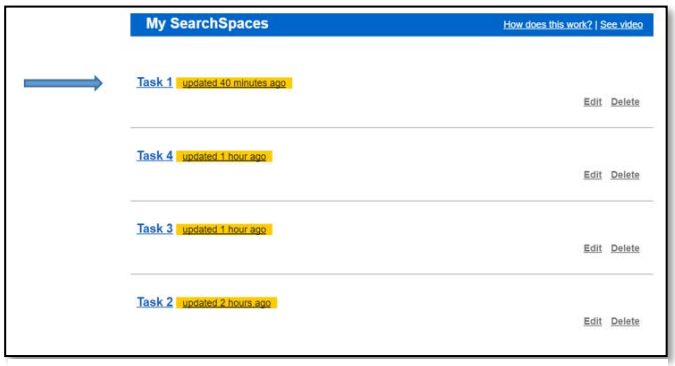


Figure 12: SearchTeam - List of Shared Folders

2. Click on **Search** tab.

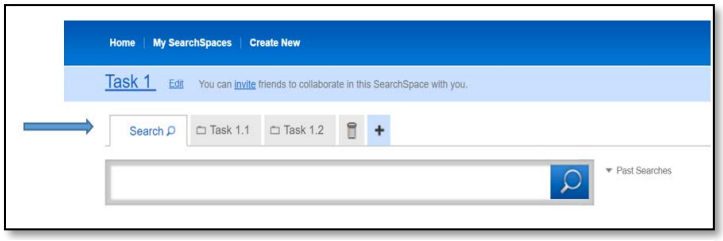


Figure 13: SearchTeam - Search Area

3. Use **Team Chat** showing on the right side to communicate with your partner during the search session and plan your search tasks.

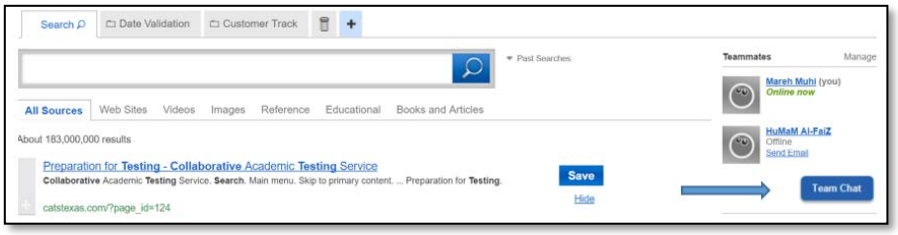


Figure 14: Search Team - Chat Tool

4. Use the search bar to insert your queries and begin the search.

5. Use the **Save** button to save results found for an assigned task. A popup will be displayed.

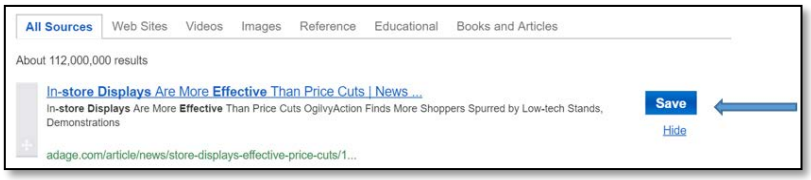


Figure 15: SearchTeam - Save Feature

6. Pick a folder to save your search results (for example, if you are working on task 1.1, pick Task 1.1 folder).

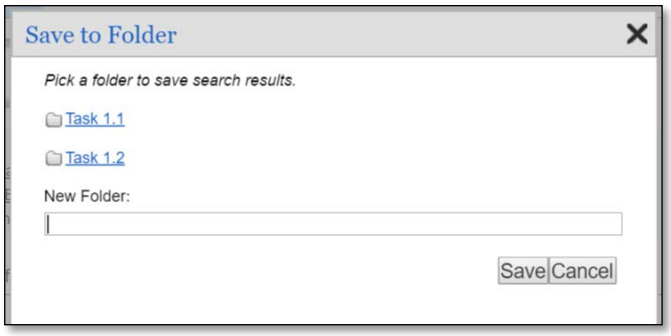


Figure 16: SearchTeam - Save to Folder Popup

7. This folder is a shared workspace, where you and your partner can add the most relevant search results.

8. Once you saved a link to a specific folder, pick that folder from the tab bar.

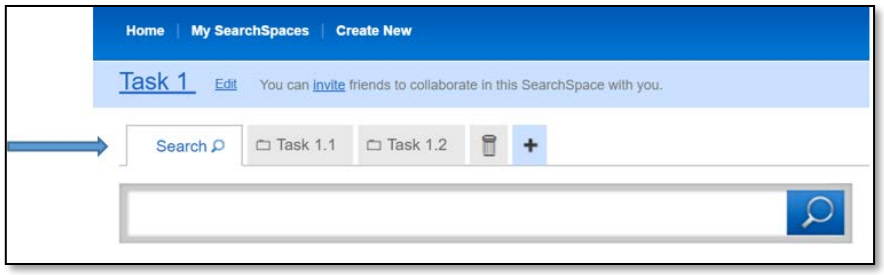


Figure 17: SearchTeam - Sub Folders Option

9. Use the **comment** area to write a short description of why you think this is a useful sit.

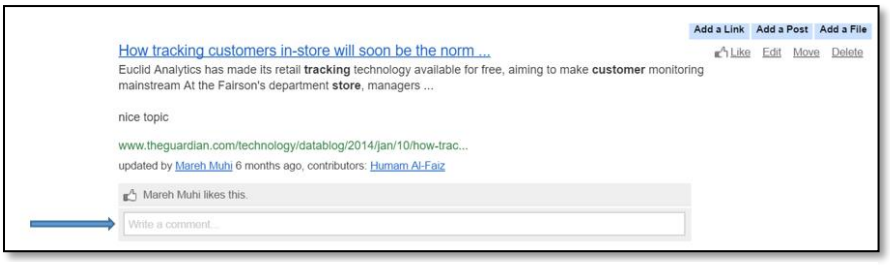


Figure 18: SearchTeam - Comment Feature

10. You can use Move button to move to another folder. You can use Delete to remove the link from a folder.

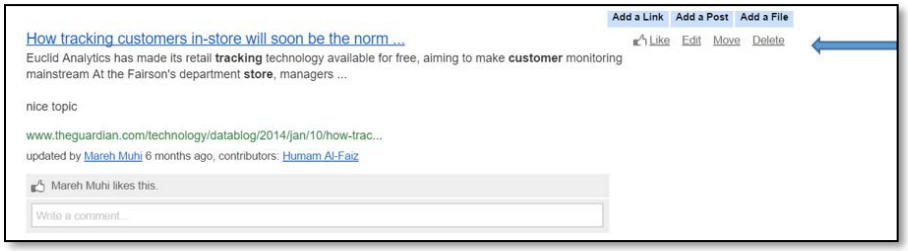


Figure 19: SearchTeam - Delete Feature

11. Repeat the same steps for the remaining search tasks.

Note: The provided folders in each search space should only contain the final results that each team have reached after working collaboratively on a task.

Saving Search Results Instructions for Google Users

We will ask the participants to work collaboratively with a partner to search the Internet for solutions to a list of software development tasks. For the convenience of the study subjects, we will create two Google accounts that will be used by participants as a part of the data collection process. Users will be using Google Drive and Google Hangouts to communicate throughout the search session.
The following list of instructions will be provided to the users:

1. Use Google Hangouts to communicate with your partner during the search session and plan your search tasks.
2. Use Google to search the Internet for a solution.
3. Use the provided file named Shared Workspace on Google Drive to save the final results. This file is shared both you and your partner can edit it at the same time.
4. Use the table structure provided to save your search results.

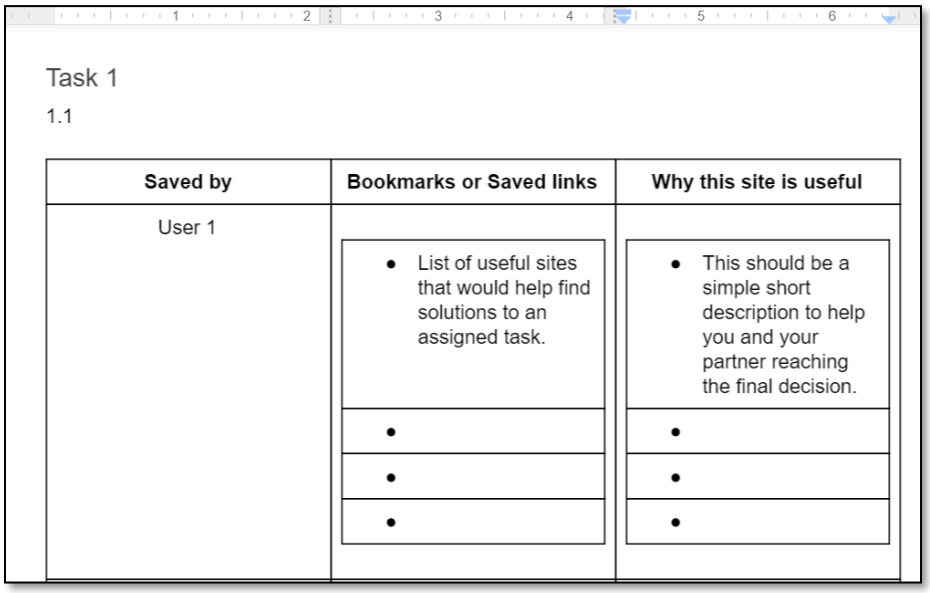


Figure 20: Google Drive - Shared Document

5. Enter your search information next to your nickname that is given to you by the investigators and listed in the **Saved by**.
6. Add your search results in **Bookmarks or Saved links**.
7. Write a short description of why you think this is a useful site use in **Why this site is useful**.

Note: The **Shared Workspace** file on **Google Drive** should only contain the final results that each team have reached after working collaboratively on a task.

APPENDIX C - POST SEARCH SURVEY

Information Seeking Questions

Rate following listed sources that influenced the way you issued queries (search terms) to identify relevant solutions for the given problems using a scale of 1 (very low influence) to 5 (very high influence)

	Very High Influence (5)	**High Influence (4)**	**Neither (3)**	**Low Influence (2)**	**Very Low Influence (1)**
Previously viewed webpages					
Previously saved webpages by yourself					
Previously saved webpages by your partner					
Previously issued queries by yourself					
Previously issued queries by your partner					
Discussion with your partner					

What was your approach or strategy towards achieving your goal of searching relevant solutions?

- ☐ The divide-and-conquer strategy: Explicitly dividing the search task into sub-tasks for each team member to undertake.
- ☐ The brute force strategy: All parties searched separately and whomever comes up with an interesting result sends it to the other and results were merged afterwards.
- ☐ Other

What challenges, if any, did you encounter related to searching and managing results found?

Communication Questions

Did you ever communicate with your partner during searches?

- ☐ Yes
- ☐ No

How did you share information with your partner?

- ☐ Using IM or team chat tool only.
- ☐ Using video calls in Google Hangouts.
- ☐ Using comments in SearchSpaces (for Search Team users): An area to show the documents saved by the participants.
- ☐ Using shared document in Google Drive.
- ☐ Other

Tool Use Questions

In what ways was the search tool useful during your task? Rate the following using a scale of 1 (Not at All Useful) to 5 (Extremely Useful).

	Extremely Useful (5)	**Very Useful (4)**	**Neither (3)**	**Slightly Useful (2)**	**Not at All Useful (1)**
The ability to save pages to a joint repository and, correspondingly, to see pages saved by another collaborator.					
The ability to leave comments on saved pages.					
The ability to view past searches.					
Use the provided chat tool to facilitate division of labor and task discussions.					
Complete the search task in a short time.					

Did you use any other tools or methods to capture information?
(e.g. paper notes, spreadsheets)

- ☐ Yes, if so, what were they?
- ☐ No

End the Process Questions

How did you decide that the quantity of information you found was enough? (As in, at what point did you decide to terminate information seeking activities).

- ☐ Verified my search results using the provided source code.
- ☐ Had a discussion with a partner about the search results.
- ☐ Other

Overall, how would you describe your success in identifying solutions for the given problems using the search tool (for SearchTeam users only)?

- ☐ The tool brought structure and persistence to the collaboration process.
- ☐ The tool unusable and I prefer the ad hoc workarounds to achieve a search task collaboratively.
- ☐ Other

Overall, how would you describe your success in identifying solutions for the given problems using the search tool (for Google users only)?

- ☐ The tool brought structure and persistence to the collaboration process.
- ☐ Way too much overhead and I prefer to use specific tool that is designed to support collaborative search activities.
- ☐ Other

Is there anything else you would like to add that we have not covered?

APPENDIX D - IRB DOCUMENTS

IRB Approval Letter

UNF
UNIVERSITY of NORTH FLORIDA.

Office of Research and Sponsored Programs
1 UNF Drive
Jacksonville, FL 32224-2665
904-620-2455 FAX 904-620-2457
Equal Opportunity/Equal Access/Affirmative Action Institution

MEMORANDUM

DATE: November 1, 2016

UNF IRB Number: 962132-2
Approval Date: 11-1-2016
Expiration Date: 11-1-2017
Processed on behalf of UNF's IRB

TO: Dr. Karthikeyan Umapathy
Computing

FROM: Dr. Jennifer Wesely, Chairperson
On behalf of the UNF Institutional Review Board

RE: Review of New Project by the UNF Institutional Review Board
IRB#962132-2: "Impact of a Collaborative Web Search Tool on Novice's Query Behavior - empirical study"

This is to advise you that your project, "Impact of a Collaborative Web Search Tool on Novice's Query Behavior - empirical study," underwent "Expedited" Categories 6 & 7 review on behalf of the UNF Institutional Review Board. Your reviewer recommended approval without further modifications.

This approval applies to your project in the form and content as submitted to the IRB for review. All participants must receive a stamped and dated copy of the approved informed consent document when possible. Any variations or modifications to the approved procedures or documents must be cleared with the IRB prior to implementing such changes. For example, if you plan to make changes to your stamped and dated informed consent form, it will be necessary to submit a copy of the revised form via an amendment so that it can be reviewed and approved prior to use. Once approved, a new stamp and date will be included on the revised consent form so that it can be used. To submit an amendment, please complete an Amendment Request Document and submit it along with any updated documents affected by the changes via a new package in IRBNet. Any unanticipated problems involving risk and any occurrence of serious harm to subjects and others shall be reported by completing this Event Report Form and sending it promptly to the IRB within 3 business days.

Your study has been approved for a period of 12 months as of 11/1/2016. If you would like your project to continue for more than one year, you will be required to provide a completed Status Report and other continuing review documentation to the UNF IRB prior to **10/1/2017.** An extension will be necessary if your study will be continuing past the 1-year anniversary of the approval date. *We ask that you submit your status report and other continuing review information 30 days before the expiration date as noted above to allow time for review and processing.* When you are ready to close your project, please complete a Closing Report Form. Please note that it will be necessary to create a new package in IRBNet in order to submit amendments, status reports, or closing reports in the future. All applicable records relating to this research shall be retained for at least 3 years after completion of the research.

CITI Training for this Project:

Name	CITI Expiration Date
Dr. Karthikeyan Umapathy	7/12/2018
Ms. Mareh Al-Sammarraie	7/9/2019

CITI Course Completion Reports are valid for 3 years. The CITI training for renewal will become available 90 days before the current CITI training expires. Please renew your CITI training when necessary and ensure that all key personnel maintain current CITI training. Individuals can access CITI by following this link: http://www.citiprogram.org/. Should you have questions regarding your project or any other IRB issues, please contact the research integrity unit of the Office of Research and Sponsored Programs by emailing IRB@unf.edu or calling (904) 620-2455.

This letter has been electronically signed in accordance with all applicable regulations, and a copy is retained within UNF's records. All records shall be accessible for inspection and copying by authorized representatives of the department or agency at reasonable times and in a reasonable manner. A copy of this approval may also be sent to the dean and/or chair of your department.

UNF IRB Number: 962132-2
Approval Date: 11-1-2016
Expiration Date: 11-1-2017
Processed on behalf of UNF's IRB

APPENDIX E – SURVEY REPORT

Default Report
Collaborative Web Search - Post Search Questionnaire

Q1 - Rate following listed sources that influenced the way you issued queries (search terms) to identify relevant solutions for the given problems using a scale of 1 (very low influence) to 5 (very high influence)

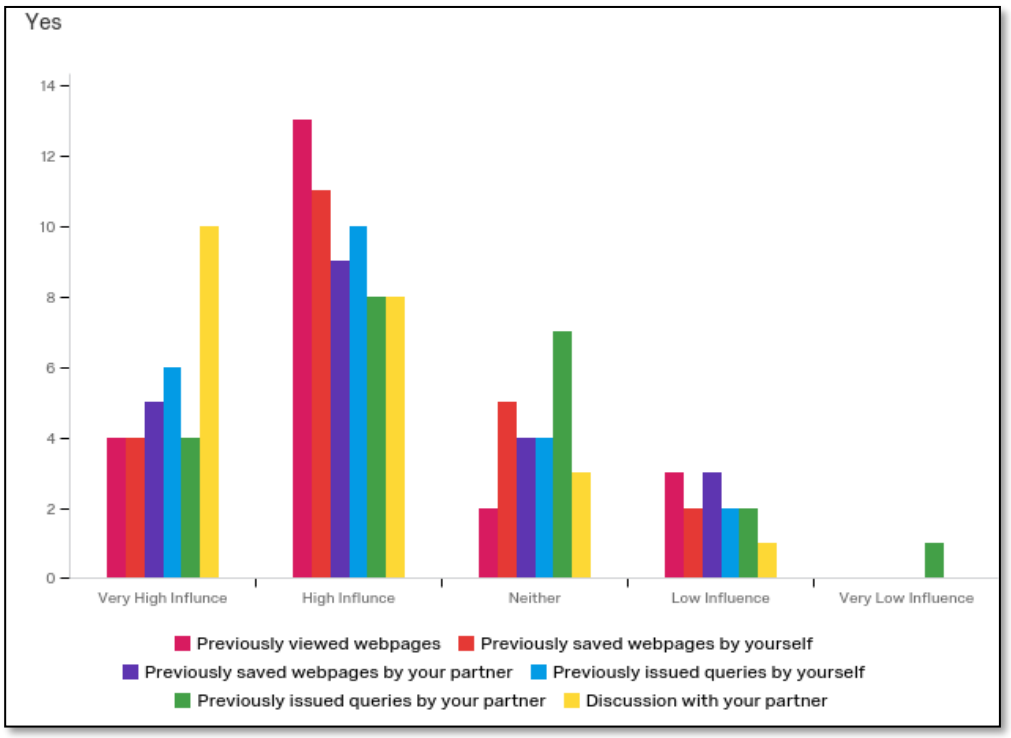


Figure 21: Participants' Responses to Q1

Q2 - What was your approach or strategy towards achieving your goal of searching relevant solutions?

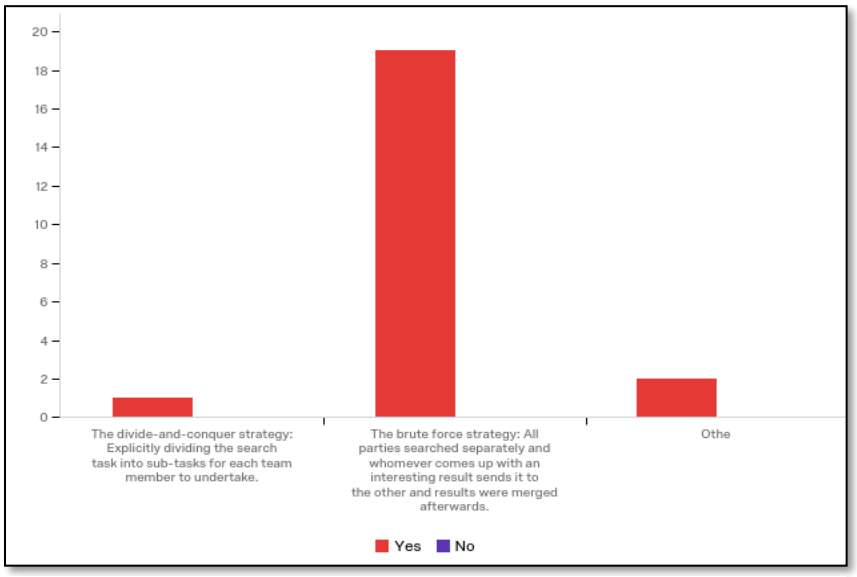


Figure 22: Participants' Responses to Q2

Other
I allowed her to search for the solution on her own unless I felt it necessary to push her in a specific direction
Break down the task in complexity and search by key words

Table 17: Other Responses to Q2

Q3 - What challenges, if any, did you encounter related to searching and managing results found?

NONE
Shot time
I was able to find answers to the given task, however I didn't always understand what I found.
We weren't able to cover every aspect of the required task, but we managed to get most of it.

I forgot to switch to a different task to get to a different chatroom
the search results may not have been up-to-date (i.e.-- some of the search results were showing for MVC 3, 4 etc) and may have been outdated answers
Sometimes my preconceived notions influenced what I was searching instead of maybe looking at other ways of doing it
some the search results were not really specific, but rather general
Going through the information that the search brought back was a little bit of a challenge.
It would be nice if the chat option popped up as a modal window regardless of what window you were currently looking at on the screen.
no history of searches unless saved - thats a bummer
I wasn't sure if the solutions given to us worked
none that are out of the ordinary. Search results were pretty decent
There were multiple solutions to approach an answer
Trying to word my searches like I would if I was looking for similar solutions in the languages I commonly use.
understanding what is relevant and what isn't if you're new to programming
not knowing exactly the terms I needed to look up
Searching and managing search results: None
the code was not very clear

Table 18: Participants' Responses to Q3

Q4 - Did you ever communicate with your partner during searches?

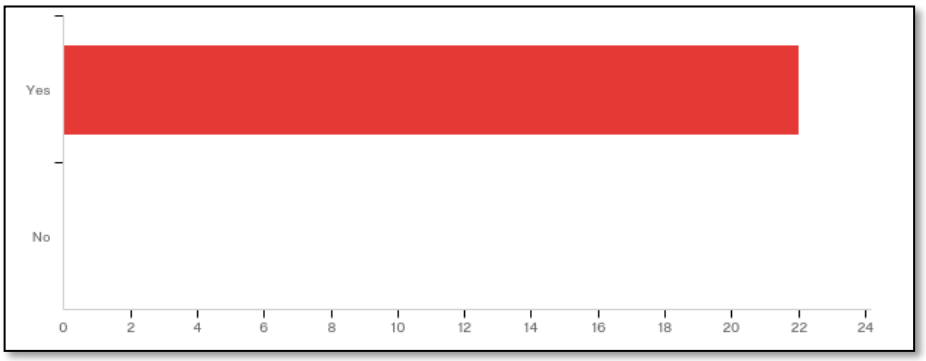


Figure 23: Participants' Responses to Q4

Q5 - Which communication tool did you use mostly to share information with your partner?

#	Question	Yes		No	
1	Using IM or team chat tool only.	66.67%	18	0.00%	0
2	Using video calls in Google Hangouts.	11.11%	3	0.00%	0
3	Using comments feature in SearchSpaces (for Search Team users).	7.41%	2	0.00%	0
4	Using Google Docs or shared document in Google Drive.	11.11%	3	0.00%	0
5	Other	3.70%	1	0.00%	0
	Total	Total	27	Total	0

Table 19: Participants' Responses to Q5

Q6 - In what ways was the search tool useful during your task? Rate the following using a scale of 1 (Extremely useless) to 5 (Extremely useful)

#	Question	Extremely Useful		Moderately Useful		Neither Useful nor Useless		Moderately Useless		Extremely Useless	
1	The ability to save pages to a joint repository and, correspondingly, to see pages saved by another collaborator.	22.41%	13	16.13%	5	15.79%	3	0.00%	0	100.00%	1
2	The ability to leave comments on saved pages.	13.79%	8	12.90%	4	52.63%	10	0.00%	0	0.00%	0
3	The ability to view past searches.	18.97%	11	16.13%	5	31.58%	6	0.00%	0	0.00%	0
4	Use the provided chat tool to facilitate	22.41%	13	25.81%	8	0.00%	0	100.00%	1	0.00%	0

	the division of labor and task discussions.										
5	Complete the search task in a short time.	22.41%	13	29.03%	9	0.00%	0	0.00%	0	0.00%	0
	Total	Total	58	Total	31	Total	19	Total	1	Total	1

Table 20: Participants' Responses to Q6

Q7 - Did you use any other tools or methods to capture information? (e.g. paper notes, spreadsheets)

#	Question	Yes		No	
1	Yes, if so, what were they?	4.55%	1	0.00%	0
2	No	95.45%	21	0.00%	0
	Total	Total	22	Total	0
Yes, if so, what were they?					
Notepad					

Table 21: Participants' Responses to Q7

Q8 - How did you decide that the quantity of information you found was enough? (As in, at what point did you decide to terminate information seeking activities).

#	Question	Yes		No	
1	Verified my search results using the provided source code	9.09%	2	0.00%	0
2	Had a discussion with a partner about the search results.	72.73%	16	0.00%	0
3	Other	18.18%	4	0.00%	0
	Total	Total	22	Total	0

Table 22: Participants' Responses to Q8

Other
my search results started circling back to some of the previous searches and the information I was finding was providing the same results
Based on prior experience and a cursory glance at the pages, I was able to determine if the solution would work or not
I ended up having to just guess
Prior work experience with related tasks

Table 23: Participant's Other Responses for Q8

Q9 - (For SearchTeam users only). Overall, how would you describe your success in identifying solutions for the given problems using the search tool?

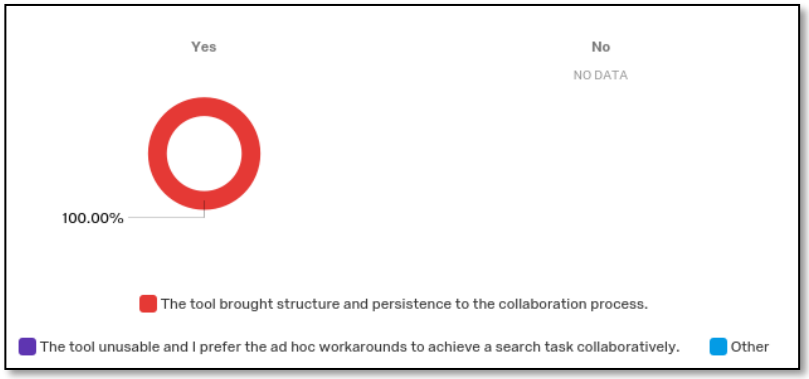


Figure 24: Participants' Responses to Q9

#	Question	Yes		No	
1	The tool brought structure and persistence to the collaboration process.	100.00%	14	0.00%	0
2	The tool unusable and I prefer the ad hoc workarounds to achieve a search task collaboratively.	0.00%	0	0.00%	0
3	Other	0.00%	0	0.00%	0
	Total	Total	14	Total	0

Table 24: Participants' Responses to Q9

Q10 - (For Google Integrated Platforms users only). Overall, how would you describe your success in identifying solutions for the given problems using the search tool?

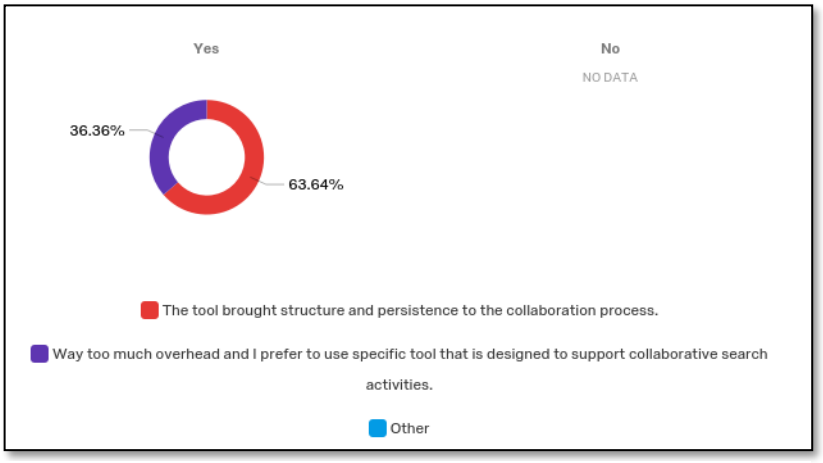


Figure 25: Participants' Responses to Q10

#	Question	Yes		No	
1	The tool brought structure and persistence to the collaboration process.	63.64%	7	0.00%	0
2	Way too much overhead and I prefer to use specific tool that is designed to support collaborative search activities.	36.36%	4	0.00%	0
3	Other	0.00%	0	0.00%	0
	Total	Total	11	Total	0

Table 25: Participants' Responses to Q10

Q11 - Is there anything else you would like to add that we have not covered?

Is there anything else you would like to add that we have not covered?
NO
Nope, that's it.
Nope. This was actually kind of fun. Good job!
sometimes I wasn't sure if partner was stuck or what the progress was and found myself waiting for messages in the chat.
Keep this short and simple and will do just fine.
none

You spelled influence wrong at the top of this survey - I enjoyed this exercise very interesting
This was really cool! :)
n/a
I hope I wasn't overbearing with the student, but I knew about these tasks and what search results would help.

Table 26: Participants' Responses to Q11

VITA

Mareh Al-Sammarraie has a Bachelor of Science degree from Jordan in Software Engineering and expects to receive a Master of Science in Software Engineering from the University of North Florida in August 2017. Mareh is currently employed as a junior Software Engineer at financial services company in Jacksonville. Mareh has been working as a software engineer for 2 years.

Mareh specializes in web development and has a strong interest in developing easy to use web applications. Mareh has professional experience developing within the Microsoft .NET technology stack, and with common web technologies like JavaScript and CSS. For more information, you can visit Mareh's LinkedIn profile here: https://www.linkedin.com/in/mareh-alsammarraie/.