May 2022

Application of the Markov Chain Method in a Health Portal Recommendation System

Xin Cai
University of Wisconsin-Milwaukee

APPLICATION OF THE MARKOV CHAIN METHOD IN A HEALTH PORTAL RECOMMENDATION SYSTEM

by

Xin Cai

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Information Studies

at

The University of Wisconsin-Milwaukee

May 2022

ABSTRACT

APPLICATION OF THE MARKOV CHAIN ANALYSIS METHOD IN A HEALTH PORTAL RECOMMENDATION SYSTEM
by

Xin Cai

The University of Wisconsin-Milwaukee, 2022
Under the Supervision of Professor Jin Zhang

This study produced a recommendation system that can effectively recommend items on a health portal. Toward this aim, a transaction log that records users' traversal activities on the Medical College of Wisconsin's HealthLink, a health portal with a subject directory, was utilized and investigated. This study proposed a mixed-method that included the transaction log analysis method, the Markov chain analysis method, and the inferential analysis method. The transaction log analysis method was applied to extract users' traversal activities from the log. The Markov chain analysis method was adopted to model users' traversal activities and then generate recommendation lists for topics, articles, and Q&A items on the health portal. The inferential analysis method was applied to test whether there are any correlations between recommendation lists generated by the proposed recommendation system and recommendation lists ranked by experts. The topics selected for this study are Infections, the Heart, and Cancer. These three topics were the three most viewed topics in the portal.

The findings of this study revealed the consistency between the recommendation lists generated from the proposed system and the lists ranked by experts. At the topic level, two topic recommendation lists generated from the proposed system were consistent with the lists ranked by experts, while one topic recommendation list was *highly* consistent with the list ranked by experts. At the article level, one article recommendation list generated from the proposed system was consistent with the list ranked by experts, while 14 article recommendation lists were *highly* consistent with the lists ranked by experts. At the Q&A item level, three Q&A item recommendation lists generated from the proposed system were consistent with the lists ranked by experts, while 12 Q&A item recommendation lists were *highly* consistent with the lists ranked by experts. The findings demonstrated the significance of users' traversal data extracted from the transaction log. The methodology applied in this study proposed a systematic approach to generating the recommendation systems for other similar portals. The outcomes of this study can facilitate users' navigation, and provide a new method for building a recommendation system that recommends items at three levels: the topic level, the article level, and the Q&A item level.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

It has been a long, but exciting adventure, and I have been fortunate enough to explore how far I can push myself during this journey.

I am forever grateful to my mentor and advisor, Dr. Jin Zhang, for his passionate and unconditional support during this journey. Without his impeccable guidance and encouragement, it would have been impossible for me to complete this dissertation. Talking and discussing with him has always been enlightening and delightful. I can still remember the day I first meet with him, he not only gave me advice about the study and how to survive in academia, but also how to fit in a foreign country as an international student.

I am grateful to have exceptional committee members: Dr. Diemar Wolfram, Dr. Iris Xie, Dr. Xiangming Mu, and Dr. Kun Lu. Thank you so much for taking the precious time to read my drafts and provide me with enlightening suggestions and feedback.

Last but not least, I would like to express my deepest gratitude to my family and friends. In particular, my wife Liqian Bao, thank you for being there throughout the ups and downs of my long journey, and my newborn son, Huanxituo, for being so adorable that I can forget anything unhappy by seeing your laugh.

Chapter 1. Introduction

1.1 Background

The Internet has grown exponentially over the last few decades and information overload has become a primary obstacle for users in finding relevant information (Eppler & Mengis, 2004). There are over 1 billion websites on the Internet and over 3 billion Internet users across the world (Internet live stats, 2016). As the Internet has grown, so too have online information sources. Consumers can access diverse online health information resources. Morahan-Martin (2008) revealed that most consumers seek health-related information from several sources: health websites, government-owned websites, online drug reference guides, medical search engines, mail lists, online social support groups, online ads, and so on.

The online health portal is one of the most significant online venues for the seeking of credible health information. It is a system that usually includes a subject directory, search engine, personalized system, and an online user community (Luo & Najdwai, 2004). Additionally, health portals are systems with complicated features and abundant information. While health portal users might have difficulties in finding needed information, two major information retrieval mechanisms have been designed to help users find information. Query-based search engines meet users' needs by returning relevant information based on users' queries. Examples include Google and Bing. The other one is the subject directory, providing a hierarchical structure for users to navigate the system.

These two information retrieval mechanisms greatly help users in many ways. However, these techniques require users to possess searching experience as well as health-related knowledge. The better the user's search strategies, the better the results returned. To improve users' experience, recommendation systems can be incorporated into the health portals for recommending related information to them. The recommendation system provides a list of items to a user, and the user selects one or more items to find out more information (Pazzani & Billsus, 2007). These systems usually adopt two types of methods: content-based methods and collaborative filtering methods. Content-based systems recommend items based on the features of the items, whereas collaborative filtering systems recommend items based on the characteristics of the users themselves.

This study develops a recommendation system for the HealthLink portal. It utilizes the transaction log to understand users' traversal patterns. The transaction log records users' activities, which include visiting an article, downloading an image, searching a term, and so on.

The Markov chain analysis method is applied to develop the recommendation system. The HealthLink portal offers dynamic user services. Users can jump from one article to another, or between topics on the portal. These kinds of activities are directional. Therefore, a traversal path for each user can be generated from a transaction log. The articles and topics can be represented by the states in a Markov chain process. In addition, the numbers of articles and topics on the portal are finite. It means that the states of the Markov chain process are also finite, which satisfies another requirement of the Markov chain method, the states should be finite or countably infinite. Moreover, the Markov chain analysis method offers a mechanism to predict future states based on the previous data. This mechanism is suitable for building a recommendation system for the health portal. There are two major advantages of using the Markov chain analysis. Firstly, the recommendation system designed based on the Markov chain theory can provide dynamic recommendation results. The traditional recommendation based on the content similarity between two items can only offer a fixed recommendation list if the items in a system remain the same. However, with the help of the Markov chain analysis method, the newly generated users' traversal paths extracted from the recent transaction logs can have an impact on the recommendation results. Moreover, the transition matrix can be modified based on the traversal path from a specific user. It means that the recommendation system based on the Markov chain theory can offer user-based recommendation.

The Markov chain analysis method, the transaction log analysis method and the inferential analysis method are proposed to conduct a systematic analysis method to recommend relevant information at different levels - the topic level, the article level, and the Q&A item level - for the HealthLink portal users.

1.2 Research problem, questions and hypotheses

The health portal has become one of the most crucial online venues for the seeking of health-related information and professional medical advice. Previous studies have focused on investigating the user's information behaviors and how to improve users' experiences. Nonetheless, few studies examined how to recommend related and relevant information for users while navigating the health portals.

1.2.1 Research problem

The primary research problem of this study is to investigate whether traversal activities from the HealthLink transaction log can be used to effectively recommend related and relevant information for users during the Healthlink portal navigation. The HealthLink portal provides consumer health information services to its patients and the community. It incorporates a wide range of health-related articles which include regular articles and Q&A items. The regular articles are published by health professionals on certain topics, whereas the Q&A items aim to answer general questions brought up by users. All users' activities are recorded in transaction logs, which include visiting an article, downloading an image, searching a term, and so on. This study centers on investigating users' traversal activities, which indicate how users navigate on the portal.

The primary goal is to see whether the recommended information generated by the Markov Chain method based on users' traversal activities on the HealthLink portal is effective. Derived from the primary problem, this study explores three research questions and the related sub-questions. To answer the research questions, the corresponding null hypotheses are proposed.

1.2.2 Research question 1 (RQ1)

RQ1: Are the topic recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts?

To address the primary research problem, the first research question is designed to uncover the hidden traversal patterns between nodes in the subject directory. Many portals including health portals offer related topics to users while they browse topics; however, the HealthLink portal doesn't offer that function. Therefore, the first object of this study is to design a topic recommendation function. All articles on HealthLink are categorized into forty-seven nodes based on article topics; each node represents a topic. The traverses between nodes can be identified by examining the transaction log. If a transaction shows that a user jumps from one article to another, and these two articles belong to different topics, a traverse between two nodes can be found. If these two articles belong to the same topic, it suggests a self- traverse within the topic. A 47×47 matrix is created to represent the traverses between nodes. Each element of the matrix indicates the volume of traverses from one node to the other. The Markov Chain analysis is adopted to model traversal activities. A recommendation list that includes relevant nodes is offered to users. Experts are recruited for evaluating the recommendation list which serves as the benchmark.

1.2.2.1. RQ1.1, RQ 1.2, & RQ1.3

To answer RQ1, three related sub-question RQ1.1, RQ1.2, and RQ1.3 are proposed. There are forty-seven topic recommendation lists since the HealthLink portal contains forty-seven subject nodes. In this study, three subject nodes - *Infection*, *Heart*, and *Cancer* - are selected to test the performance of the topic recommendation system. The reasons for selecting these three topics are that they not only have significant traffic volumes on the portal but also have large article collections. Therefore, the following three sub-questions are examined.

RQ1.1: Is the topic recommendation list generated by the proposed recommendation system consistent with the list ranked by the experts in terms of the *Infection* topic?

RQ1.2: Is the topic recommendation list generated by the proposed recommendation system consistent with the list ranked by the experts in terms of the *Heart* topic?

RQ1.3: Is the topic recommendation list generated by the proposed recommendation system consistent with the list ranked by the experts in terms of the *Cancer* topic?

1.2.2.2. Hypotheses group 1

To answer the sub-questions, the inferential analysis method is applied to compare the performance of recommendation systems. Hypotheses H_{01}, H_{02}, and H_{03} will be tested to answer RQ1.1, RQ1.2, and RQ1.3 respectively:

H_{01}: There is no significant correlation between the topic recommendation list ranked by the experts and the list generated by the proposed recommendation system in terms of the Infection topic.

H_{02}: There is no significant correlation between the topic recommendation list ranked by the experts and the list generated by the proposed recommendation system in terms of the Heart topic.

H_{03}: There is no significant correlation between the topic recommendation list ranked by the experts and the lists generated by the proposed recommendation system in terms of the Cancer topic.

For H_{01}, H_{02}, and H_{03}, the two variables for the correlation tests are the topic recommendation list generated by the proposed system and the list ranked by the experts.

1.2.3 Research question 2 (RQ2)

RQ2: Are the article recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts?

The second research question aims to identify the underlying traversal patterns when users are traversing between articles on the portal. There are 2377 articles under forty-seven different topics. Since the article recommendation lists are generated based on the existing viewable articles, it's vital to select appropriate articles to test the article recommendation system. In this study, fifteen articles from the three topics - *Infection, Heart, and Cancer* - are selected respectively. Therefore, RQ2 is divided into three sub-questions for the corresponding three topics.

1.2.3.1. RQ2.1, RQ2.2, and RQ2.3

RQ2.1: Are the article recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts in terms of the *Infection* topic?

RQ2.2: Are the article recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts in terms of the *Heart* topic?

RQ2.3: Are the article recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts in terms of the *Cancer* topic?

1.2.3.2. Hypotheses group 2

Hypotheses H_{04}, H_{05}, and H_{06} are proposed to answer RQ2.1, RQ2.2, and RQ2.3 respectively:

H_{04}: There are no significant correlations between the article recommendation lists ranked by the experts and the lists generated by the proposed recommendation system in terms of the Infection topic.

H_{05}: There are no significant correlations between the article recommendation lists ranked by the experts and the lists generated by the proposed recommendation system in terms of the Heart topic.

H_{06}: There are no significant correlations between the article recommendation lists ranked by the experts and the lists generated by the proposed recommendation system in terms of the Cancer topic.

For H_{04}, H_{05}, and H_{06}, the two variables for the correlation tests are the article recommendation lists generated by the proposed system and the lists ranked by the experts.

1.2.4 Research question 3 (RQ3)

RQ3: Are the Q&A item recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts?

RQ2 answers the performance of the proposed system to recommend articles to users in general. RQ3 aims to investigate the traversal patterns among Q&A items. Q&A items make up a unique component of the portal. These items were written by health professionals for answering questions published by users. Similar to RQ2, 15 Q&A items from 3 topics are selected to test the performance of the proposed recommendation system. RQ3 is divided into the following three sub-questions for the corresponding three topics.

1.2.4.1. RQ3.1, RQ3.2, and RQ3.3

RQ3.1: Are the Q&A item recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts in terms of the *Infection* topic?

RQ3.2: Are the Q&A item recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts in terms of the *Heart* topic?

RQ3.3: Are the Q&A item recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts in terms of the *Cancer* topic?

1.2.4.2. Hypotheses group 3

Hypotheses H_{07}, H_{08}, and H_{09} are proposed to answer RQ3.1, RQ3.2, and RQ3.3 respectively:

H_{07}: There are no significant correlations between the Q&A item recommendation lists ranked by the experts and the lists generated by the proposed recommendation system in terms of the Infection topic.

H_{08}: There are no significant correlations between the Q&A item recommendation lists ranked by the experts and the lists generated by the proposed recommendation system in terms of the Heart topic.

H_{09}: There are no significant correlations between the Q&A item recommendation lists ranked by the experts and the lists generated by the proposed recommendation system in terms of the Cancer topic.

For H_{07}, H_{08}, and H_{09}, the two variables for the correlation tests are the Q&A item recommendation lists generated by the proposed system and the lists ranked by the experts.

1.2.5 Research design

To better depict the connections among the research problem, research questions, and hypotheses of the proposal, Figure 1 is created to show relationships among them. The three research questions are proposed to answer the primary research questions. Each research question contains three sub-questions. The Markov chain analysis method and the inferential analysis method are used to answer the research questions and the related sub-questions.

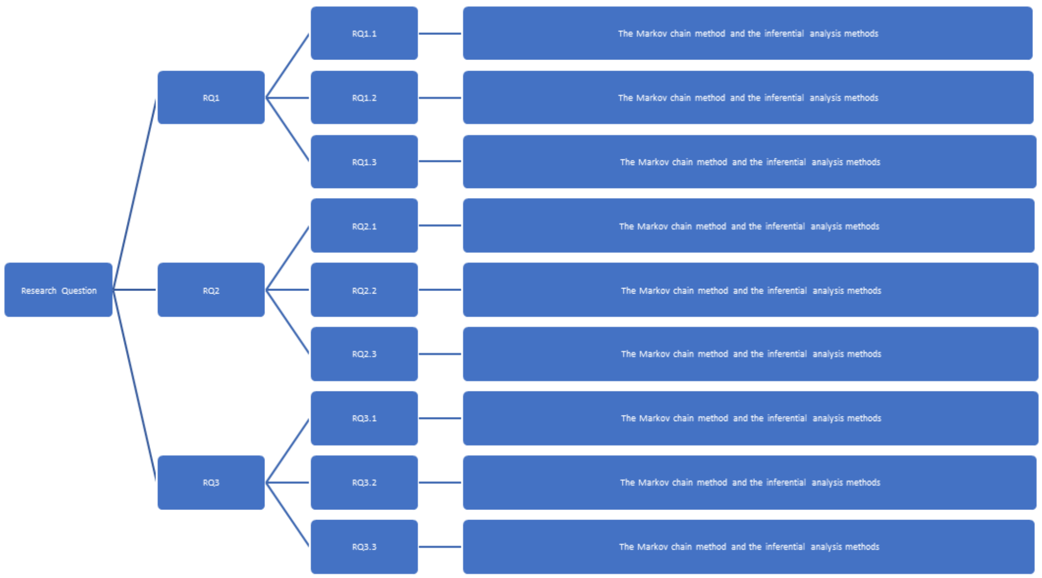


Figure 1. Research design structure

1.3 Definition of terms

1.3.1 Markov Chain

Markov chain theory, brought up by the Russian mathematician Andrey (Andrei) Andreyevich Markov (1856-1922), is a special stochastic process which is "memoryless" (Marcellin & Fischer, 1990). "Memoryless" means that the future state or outcome depends only on the information of the current time and not on the past. There are two components in the Markov chain: a set of states that contains all the possible states or variables and a set of probabilities that denotes the transition probability between two states. With these two components, the possibilities of future states can be calculated based on the current state.

1.3.2 State

In computer science and information science, states refer to the preceding or predefined events of the system. The state set contains all the possible states or variables. In this study, the state set includes three different types of states: node states (forty-seven subject directory nodes), article states (all the articles on the portal), and Q&A items states (all the Q&A items on the portal).

1.3.3 Transition matrix

The transition matrix is used to summarize the transition possibilities between states. Let S be the set of states, $S = \{\boldsymbol{s_1}, \boldsymbol{s_2}, \ldots, \boldsymbol{s_r}\}$. Let $\boldsymbol{p_{ij}}$ be the transition probability from state i to j.

Transition matrix T can be shown in Equation 1:

$$\boldsymbol{T} = \begin{pmatrix} \boldsymbol{p_{1,1}} & \cdots & \boldsymbol{p_{1,r}} \\ \vdots & \boldsymbol{p_{i,j}} & \vdots \\ \boldsymbol{p_{r,1}} & \cdots & \boldsymbol{p_{r,r}} \end{pmatrix} \quad \textbf{(1)}$$

1.3.4 Recommendation system

The recommendation system is a system that offers a list of items to a user, and the user selects one or more items to find out more information (Pazzani & Billsus, 2007). Ricci et al. (2011) put forward the definition of the item as "the general term used to denote what the system recommends to users". In this definition, the item to be recommended could be: the commodities sold in E-commerce like Amazon (Linden et al., 2003), the movies from online streaming and media service providers such as Netflix (Zhou et al., 2008) and YouTube (Davidson, 2010), the academic papers from scientific citation indexing databases (Lee et al., 2013), the webpages of Web portals (Sarukkai, 2000), and so on. In a word, from the perspectives of the information retrieval field, recommendation systems refer to the systems which satisfy information needs by suggesting items to users.

1.3.5 Health portal

The online health portal is a system that usually contains a subject directory, a search engine, a personalized system, and an online user community (Luo & Najdwai, 2004). The health subject directory facilitates Web editors to organize the health-related information and offers access like hyperlinks to the outside health information providers. The search engine function helps consumers to search for topics within the health portal or on the external websites. The personalization system provides consumers with tailored services and a mechanism to edit the user interface. The community function allows registered users and unregistered users to exchange health-related and unrelated information. These functions helped consumers search for information in more than one way. They indicate that a health portal is a combination of complicated and multifunctioning Web applications.

1.3.6 Subject directory

Subject directories are a kind of information organization system that divides information (webpages) into several subject categories based on the characteristics of webpages (Zhang et al., 2009). This system usually has a hierarchy structure since subject categories can be further divided into sub-categories. Subject directories have been adopted by numerous online portals such as Yahoo.com, health-related websites, and so on. One of the ordinary commonalities of these portals is the diversity in which they offer various types of information to users. Subject directories predominantly facilitate users' experiences on websites if they are familiar with the topic they are browsing.

1.3.7 HealthLink

The Medical College of Wisconsin's HealthLink is an online health portal that provides consumer health information services to patients and the community. Figure 2 shows its homepage on February 12, 2009. It offers two major features: "Health-Articles" and "Medical College Patient Care". Health Articles cover a variety of health-related topics. Users can find related articles in two ways: a subject directory that arranges articles based on topics and a keyword search engine. The health users can also find information using their preferred search method. HealthLink also presents patient care features including to find a doctor, find clinics and health centers, and request appointments with health providers. The HealthLink portal retired in 2009. All files of the portal were kept, which include webpages, the structure information of the portal, and the transaction log which records users' activities. The freshness of data is not the most important concern since data extracted from the HealthLink portal and the transaction log are used to test performance of the proposed recommendation system.

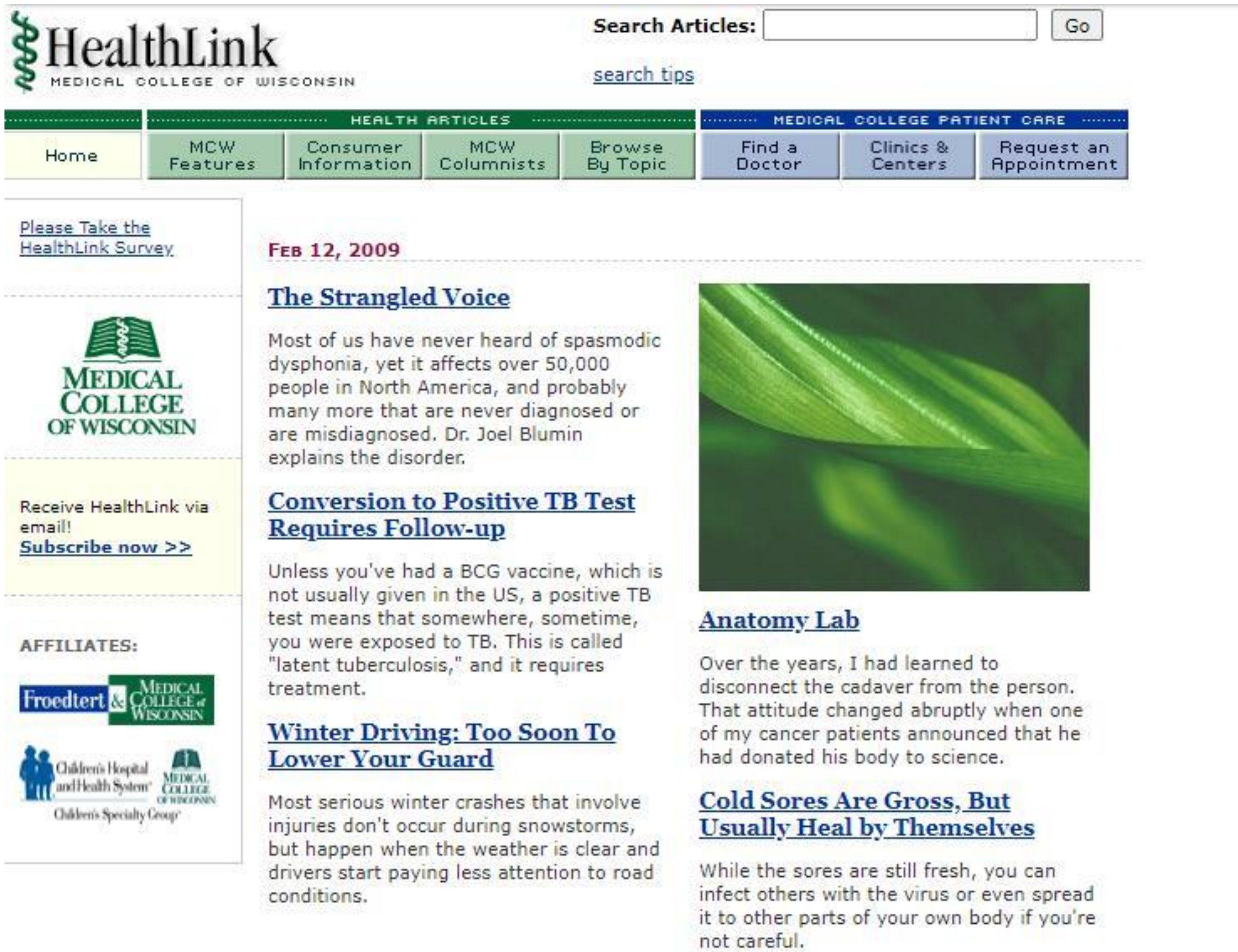


Figure 2. The homepage of the HealthLink portal

1.3.8 Topics

HealthLink divides its health-related articles into topics based on their subject characteristics. Therefore, a subject directory characterizes the system that organizes the articles. The subject directory is a three-level hierarchical system. The root is "browse by topic", where users start to find the articles. The nodes on the directory branches are topics. In the directory, there are forty-seven topics/nodes that users can find. These topics are displayed in Figure 3. Each node contains a collection of related articles and Q&A items. Users can traverse within the nodes, articles, and Q&A items at will.

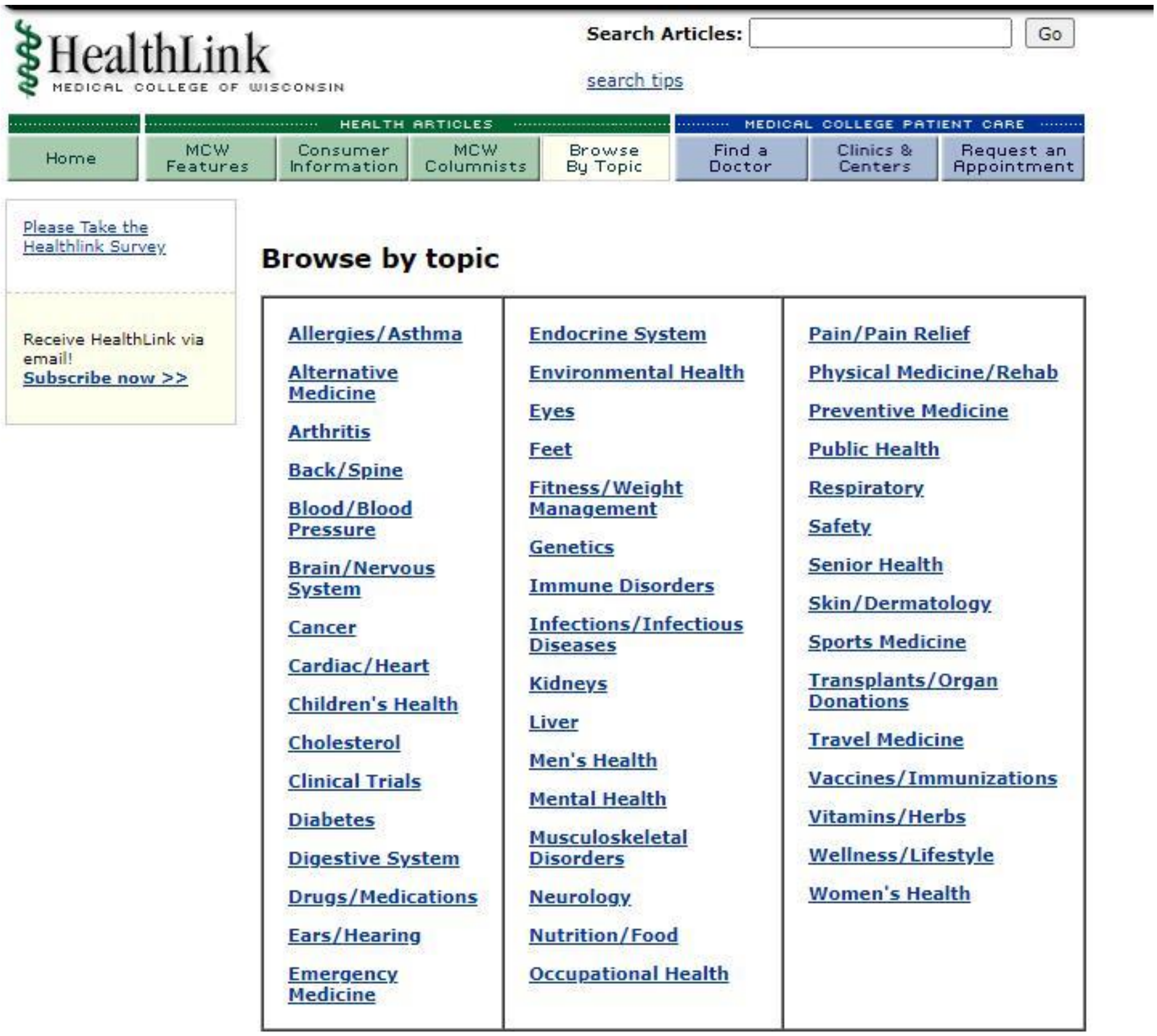
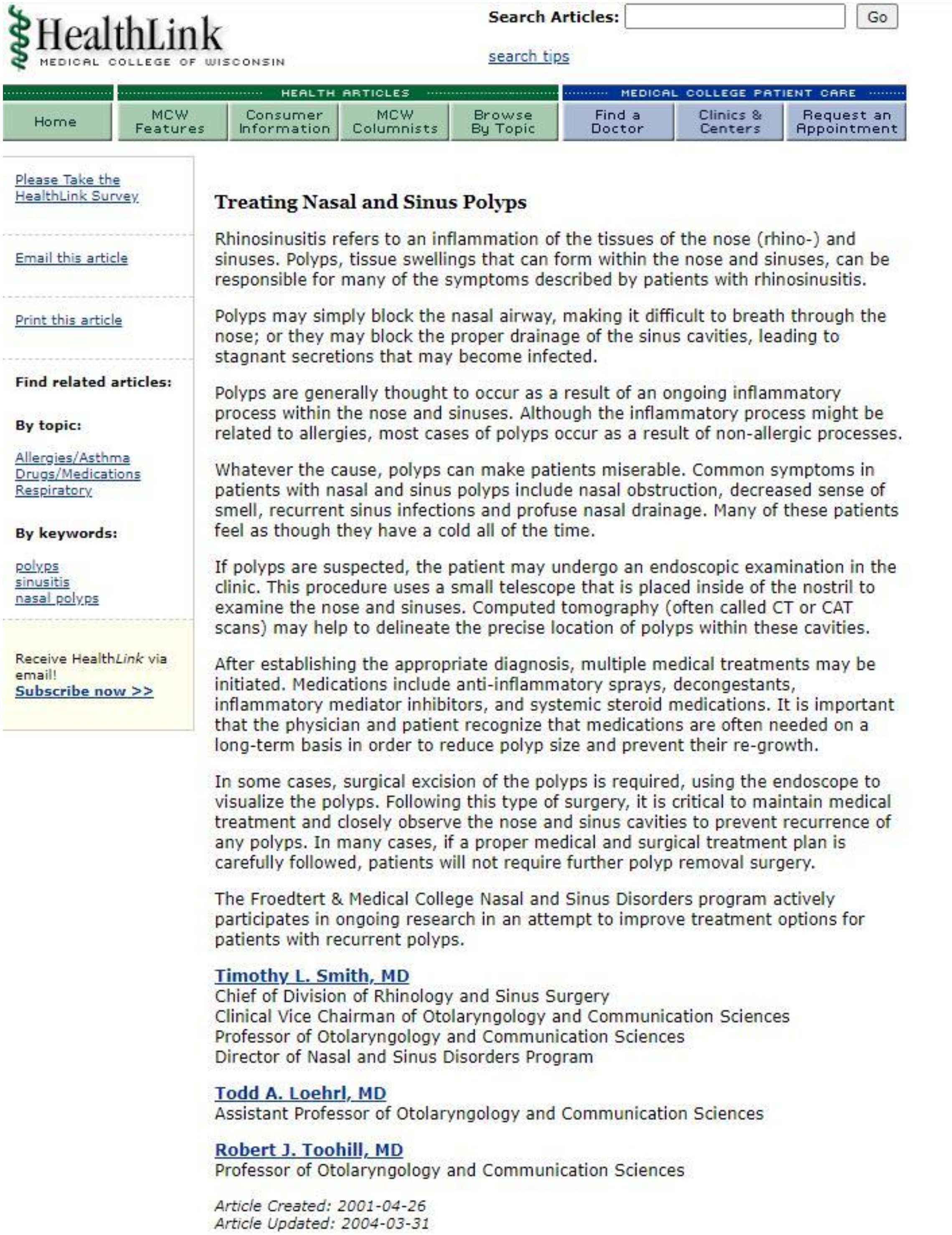


Figure 3. The page of topics root

1.3.9 Articles

HealthLink's users primarily acquire health-related information through articles. Articles come from three different sources: health professionals, website contributors, and other health-related websites and health officials. Articles usually contain titles, content, and authors' information. From the left-hand sidebar, users can take the HealthLink survey, email and print the article, and find related articles by topics or keywords. Figure 4 shows the view of an article.



HealthLink
MEDICAL COLLEGE OF WISCONSIN

Search Articles: [ ] Go

search tips

HEALTH ARTICLES | MEDICAL COLLEGE PATIENT CARE

Home | MCW Features | Consumer Information | MCW Columnists | Browse By Topic | Find a Doctor | Clinics & Centers | Request an Appointment

Please Take the HealthLink Survey

Email this article

Print this article

Find related articles:

By topic:

Allergies/Asthma
Drugs/Medications
Respiratory

By keywords:

polyps
sinusitis
nasal polyps

Receive Health*Link* via email!
Subscribe now >>

Treating Nasal and Sinus Polyps

Rhinosinusitis refers to an inflammation of the tissues of the nose (rhino-) and sinuses. Polyps, tissue swellings that can form within the nose and sinuses, can be responsible for many of the symptoms described by patients with rhinosinusitis.

Polyps may simply block the nasal airway, making it difficult to breath through the nose; or they may block the proper drainage of the sinus cavities, leading to stagnant secretions that may become infected.

Polyps are generally thought to occur as a result of an ongoing inflammatory process within the nose and sinuses. Although the inflammatory process might be related to allergies, most cases of polyps occur as a result of non-allergic processes.

Whatever the cause, polyps can make patients miserable. Common symptoms in patients with nasal and sinus polyps include nasal obstruction, decreased sense of smell, recurrent sinus infections and profuse nasal drainage. Many of these patients feel as though they have a cold all of the time.

If polyps are suspected, the patient may undergo an endoscopic examination in the clinic. This procedure uses a small telescope that is placed inside of the nostril to examine the nose and sinuses. Computed tomography (often called CT or CAT scans) may help to delineate the precise location of polyps within these cavities.

After establishing the appropriate diagnosis, multiple medical treatments may be initiated. Medications include anti-inflammatory sprays, decongestants, inflammatory mediator inhibitors, and systemic steroid medications. It is important that the physician and patient recognize that medications are often needed on a long-term basis in order to reduce polyp size and prevent their re-growth.

In some cases, surgical excision of the polyps is required, using the endoscope to visualize the polyps. Following this type of surgery, it is critical to maintain medical treatment and closely observe the nose and sinus cavities to prevent recurrence of any polyps. In many cases, if a proper medical and surgical treatment plan is carefully followed, patients will not require further polyp removal surgery.

The Froedtert & Medical College Nasal and Sinus Disorders program actively participates in ongoing research in an attempt to improve treatment options for patients with recurrent polyps.

Timothy L. Smith, MD
Chief of Division of Rhinology and Sinus Surgery
Clinical Vice Chairman of Otolaryngology and Communication Sciences
Professor of Otolaryngology and Communication Sciences
Director of Nasal and Sinus Disorders Program

Todd A. Loehrl, MD
Assistant Professor of Otolaryngology and Communication Sciences

Robert J. Toohill, MD
Professor of Otolaryngology and Communication Sciences

Article Created: 2001-04-26
Article Updated: 2004-03-31

Figure 4. The article page view

Articles are presented under the topics. Using topic 6: Cancer as an example, it contains 141 articles. A sub-tree structure is depicted in Figure 5. These 141 articles are presented on the index page of the node for diabetes. Users can easily browse through all the articles under this node if he/she is interested in this topic. The rest of the forty-six topics are all organized in the same way.

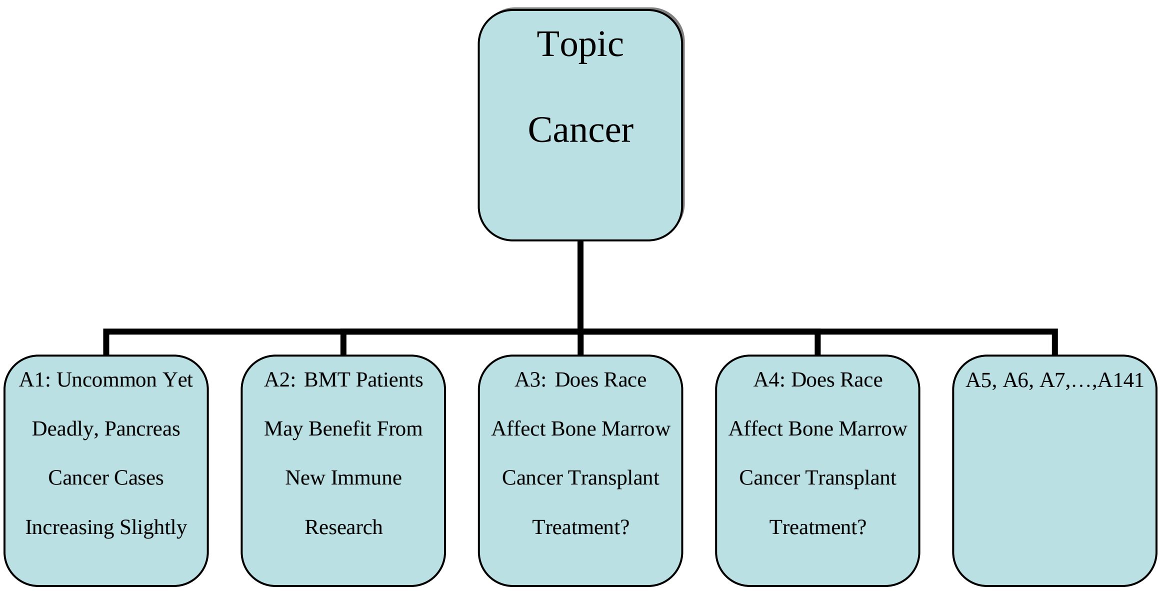
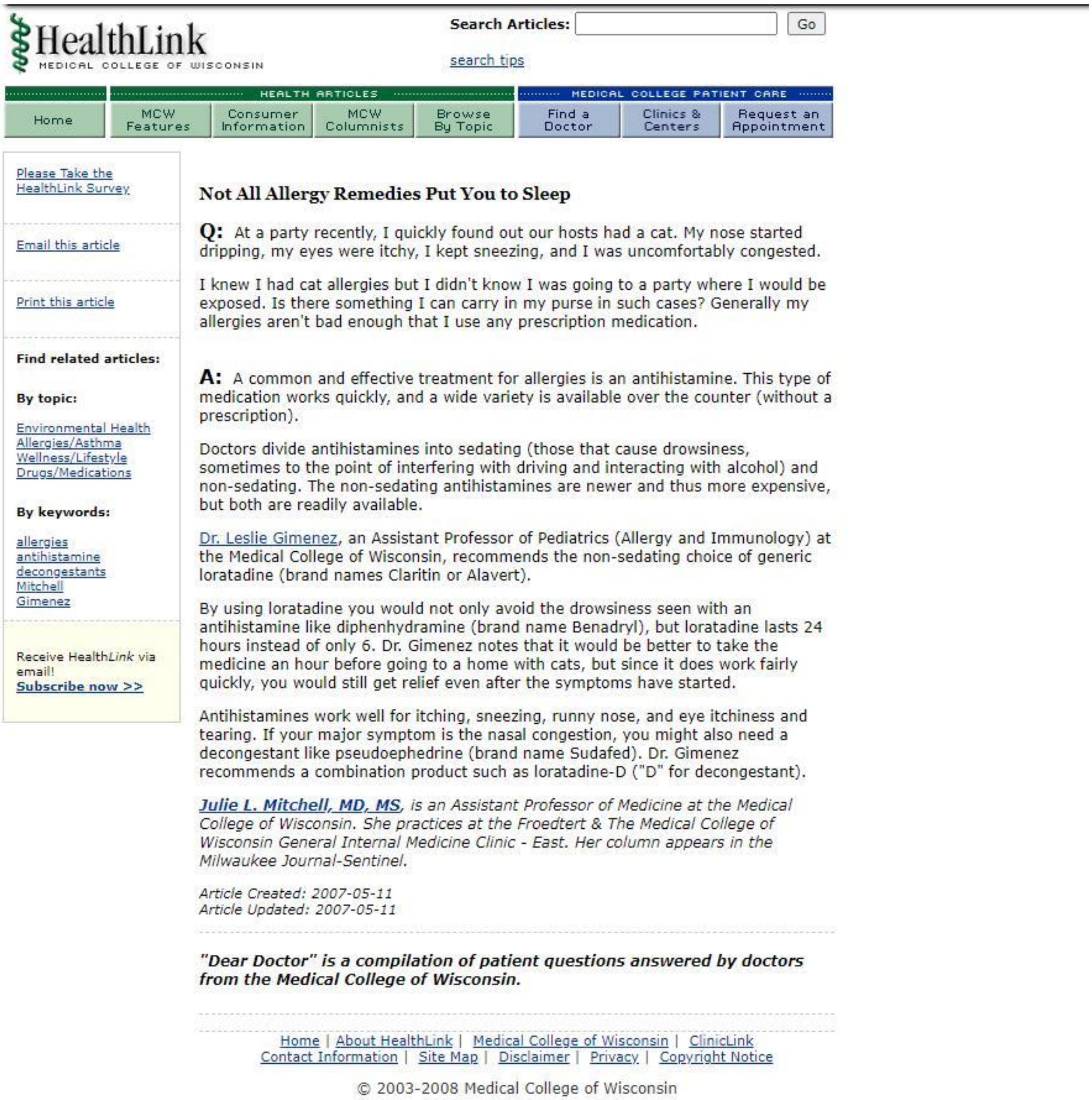


Figure 5. The Cancer topic

1.3.10 Q&A items

Although doctors and the portal's contributors have tried their best to provide timely and extensive health-related information through publishing articles, there are still various and unexpected questions asked by users. Therefore, the HealthLink portal offers a unique set of articles that answers users' issues. These are Q&A items. Editors selected the most representative questions and handed them to experienced health professionals. Q&A items are displayed like regular articles. Figure 6 shows the destination of a Q&A item, which lists questions and answers.



HealthLink
MEDICAL COLLEGE OF WISCONSIN

Search Articles: Go

search tips

HEALTH ARTICLES | MEDICAL COLLEGE PATIENT CARE

Home | MCW Features | Consumer Information | MCW Columnists | Browse By Topic | Find a Doctor | Clinics & Centers | Request an Appointment

Please Take the HealthLink Survey

Email this article

Print this article

Find related articles:

By topic:

Environmental Health
Allergies/Asthma
Wellness/Lifestyle
Drugs/Medications

By keywords:

allergies
antihistamine
decongestants
Mitchell
Gimenez

Receive Health*Link* via email!
Subscribe now >>

Not All Allergy Remedies Put You to Sleep

Q: At a party recently, I quickly found out our hosts had a cat. My nose started dripping, my eyes were itchy, I kept sneezing, and I was uncomfortably congested.

I knew I had cat allergies but I didn't know I was going to a party where I would be exposed. Is there something I can carry in my purse in such cases? Generally my allergies aren't bad enough that I use any prescription medication.

A: A common and effective treatment for allergies is an antihistamine. This type of medication works quickly, and a wide variety is available over the counter (without a prescription).

Doctors divide antihistamines into sedating (those that cause drowsiness, sometimes to the point of interfering with driving and interacting with alcohol) and non-sedating. The non-sedating antihistamines are newer and thus more expensive, but both are readily available.

Dr. Leslie Gimenez, an Assistant Professor of Pediatrics (Allergy and Immunology) at the Medical College of Wisconsin, recommends the non-sedating choice of generic loratadine (brand names Claritin or Alavert).

By using loratadine you would not only avoid the drowsiness seen with an antihistamine like diphenhydramine (brand name Benadryl), but loratadine lasts 24 hours instead of only 6. Dr. Gimenez notes that it would be better to take the medicine an hour before going to a home with cats, but since it does work fairly quickly, you would still get relief even after the symptoms have started.

Antihistamines work well for itching, sneezing, runny nose, and eye itchiness and tearing. If your major symptom is the nasal congestion, you might also need a decongestant like pseudoephedrine (brand name Sudafed). Dr. Gimenez recommends a combination product such as loratadine-D ("D" for decongestant).

***Julie L. Mitchell, MD, MS**, is an Assistant Professor of Medicine at the Medical College of Wisconsin. She practices at the Froedtert & The Medical College of Wisconsin General Internal Medicine Clinic - East. Her column appears in the Milwaukee Journal-Sentinel.*

Article Created: 2007-05-11
Article Updated: 2007-05-11

"Dear Doctor" is a compilation of patient questions answered by doctors from the Medical College of Wisconsin.

Home | About HealthLink | Medical College of Wisconsin | ClinicLink
Contact Information | Site Map | Disclaimer | Privacy | Copyright Notice

© 2003-2008 Medical College of Wisconsin

Figure 6. The Q&A item page view

1.3.11 Transaction log

Web transaction logs record the interactions between users and Web-based systems in the database. An entry in the transaction log records an action which includes an Internet Protocol (IP) address, a user machine cookie, a timestamp indicating the action time, a referrer directing to the current page, a referral which is the current page, and a session identifier which is not necessary to this study.

1.3.12 User traversal activities

While traversing within the system, users can conduct various tasks which include: traversing between articles, traversing between topics, traversing from topic page to the related articles, traversing from articles to related topics, printing the current page, emailing the current page, and searching the articles.

Traverse between articles: Users can traverse between articles by clicking embedded hyperlinks.

Traverse from topic page to the related articles: There are forty-seven topics/nodes; each topic contains articles. Users can traverse from the topic page to the related articles.

Traverse from articles to related topics: Users can jump back to topics to find more related articles.

Print the current page: Users can print the current page.

Email the current page: Users can email the current page to anyone who owns an email address or to themselves.

Search articles: Users can conduct keyword searches within the portal.

Note that there is no way that a user can traverse between topics directly in the system. A traverse between topics is defined as if there is a traverse between two articles and the corresponding topics are identified to represent the traverse

1.3.13 Sessions

A session can be viewed as a set of activities that completed by a user in a predefined period. Identifying sessions help researchers to extract activities of each user. In this study, a session is defined as all of the activities conducted by a user with a unique IP address in two hours.

1.3.14 Traversal paths

In this study, identifying traversal paths is crucial since it is fundamental to creating transition matrices. A traversal path consists of the pages a user visited during a session. Only topic pages, article pages, and Q&A item pages are taken into consideration since the major goal of this study is to understand how users traverse between topics and articles.

Chapter 2. Literature review

2.1. Introduction

This chapter reviews related studies which include the Markov chain analysis, transaction log analysis, consumer health information, online information resources, subject directories, and recommendation systems.

2.2. Markov chain in information retrieval studies

The Markov chain theory has been widely accepted in the information science field since Claude Shannon brought the thought of it in her famous publication, *A mathematical theory of communication*, for explaining the concept of “information entropy” in 1948. The Markov chain method possesses a predictive power which only requires limited pre-existing information for making the predictions.

Many information retrieval systems have taken advantage of it for improving performances. Applications include book circulation prediction in the library systems (Coady, 1983), link and path prediction in the subject directory portals (Sarukkai, 2000; Zhu, Hong, & Hughes, 2002; Eirinaki, Vazirgiannis, & Kapogiannis, 2005); user action prediction during information searching process (Benoît, 2005; Xie & Joo, 2010) and so on. Besides the field of information science, many other fields have successfully adopted the principles of the Markov chain theory, such as pattern recognition from machine learning (Bishop, 2006), protein modeling from biology (Krogh et al., 1994), modeling economic growth (Quah, 1993), and the list goes on.

The application of Markov chain method to information retrieval can be classified into two major categories: user-oriented research designs and system-oriented research designs.

2.2.1 User-oriented information retrieval studies

In the user-oriented information retrieval studies, the major objectives are identifying the user behavior patterns and building an information retrieval paradigm. The discovery of user behavior patterns is fundamental for designing systems since it helps system engineers and designers better understand what the needs of users are and how users interact with the systems and other users.

Qiu (1993) studied search patterns across different user groups in an experimental hypertext information retrieval system. The search patterns were compared to the corresponding transition matrices. The Markov models were tested by using a log-linear model. The results showed that the second-order Markov model (suggesting the previous two states influence the current state) was the best model and that significant differences exist between the studied groups in terms of users' search behavior.

Xie and Joo (2010) investigated search tactic transitions during the Web-based search process. Thirteen search tactics were identified, including determining the initial websites for starting the search, forming the search queries, modifying or extending the search queries, evaluating the relevance and quality of an individual item returned, evaluating the full list of returned items, remembering or writing down the metadata of a returned result before getting it, accessing the returned records, moving back to the previous pages, learning the system-related knowledge and domain knowledge, exploring information and knowledge of the returned results in the website, putting the results with similar characteristics in a list, monitoring the search process, and finally fulfilling information needs by using the information searched. Then the probabilities of tactic transitions were estimated by a fifth-order Markov chain method, and a transition matrix was generated. The results revealed the most common search transition patterns in the search process and suggested that the information retrieval systems should assist the users by reducing needless transitions.

Benevenuto et al. (2009) examined the information searching behaviors in online social networks using the Markov chain. The authors identified the search activities in social networks including: searching user profiles, communities, and topics; browsing and creating users' introductions; sending messages; leaving or reading comments; and uploading or watching videos, photos and community activities. An activities transition matrix was constructed from clickstream data. The probabilities of transition between search activities suggested how users interacted with friends and strangers in the social networks.

2.2.2 System-oriented information retrieval studies

The motivations of user-oriented study are to enhance the performance of information retrieval systems by better understanding users' information needs. The system-oriented information retrieval studies attempt to enrich the field by building a more advanced system, by improving or even inventing information retrieval-related algorithms, by designing a user-friendly interface and so on.

2.2.2.1. Modeling navigation pattern and link prediction

Modeling users' navigation patterns is a hot topic in system-oriented information retrieval studies (Sarukkai, 2000; Zhu et al., 2002; Sen & Hansen, 2003). In a Web portal or subject directory, the system's performance is determined by not only the quality of its contents but also the rationality and usability of the website's organization.

In addition to recommendation systems, many Web portals provide a navigational agent for Webpage recommendation when users are reading a webpage. This type of recommendation, called link prediction, helps users skip unnecessary actions (Perkowitz & Etzioni, 1997). Briefly stated, a well-functioned Web portal provides a well-organized structure that facilitates users to find relevant and useful information and even predicts the potential pages which users might want to see next.

In view of that, it is essential to study the user Web navigation behavior and the patterns for helping the Web portals in website re-organization and link recommendation. Sarukkai (2000) modeled the navigation behavior on a Web server. The traversal paths were extracted from the server's transaction log which recorded the interactions between users and the server. The author modeled the navigation patterns from the paths as a transition matrix. The states he defined were URLs, HTTP requests, and actions such as updating the database or sending an e-mail. The results indicated that the Markov chain method was promising for link recommendation and pattern modeling. Sen and Hansen (2003) put forward a more complicated model that generated link predictions by considering the current page and the last visited page. The authors compared the performances of the first-order Markov chain and the second-order Markov chain. It turned out that the second-order model worked well although it took enormous computational power. A mixed model that combined several first-order models was proposed to cluster web pages on the website. The results showed the proposed model performed well in predicting the next page while reducing the computational cost.

In order to achieve higher prediction accuracy, Eirinaki et al. (2005) noted that it was insufficient to make link suggestions which only take webpage usage data into consideration. They incorporated a PageRank-like algorithm with the Markov chain models. The results showed that the adjusted suggestions were much better in link prediction. The clustering methods have also been integrated. Borges and Levene (2004) tried to cluster users with shared interests with similar navigation patterns. They also gathered webpages with similar visiting probabilities at the same time. Cadez et al. (2000) adopted information visualization techniques to exhibit the recommended path extracted from navigation patterns.

A problem with the Markov chain-based methods is the computational cost. The dimension of the transition matrix can be colossal if a Web portal contains a tremendous number of websites like Yahoo, AOL, and MSN. With respect to that, researchers introduced dimensionality algorithms to tackle this problem which haunted the information retrieval field many years. Zhu, Hong, and Hughes (2002) applied a matrix compression algorithm which was invented by Spears (1998). The results showed that the dimensions have been reduced dramatically without significant errors. Since the transition matrix was square, Sarukkai (2000) adopted eigendecomposition in compressing the transition matrix. The eigendecomposition is a mathematical method used to factorize a matrix into a canonical form.

Besides navigation behaviors, Benoît (2005) viewed the information retrieval process as a Markov chain, and the terms used during the search were regarded as states. Construction of the transition matrix was based on the data from experienced users. The newly built system provided a dynamic and interactive information searching environment that helped the users understand the relationships between concepts by integrating visualization techniques.

2.2.2.2. Document ranking

Thanks to the dramatic development of information and communication technologies, the problem of information overload has become a critical issue for information users (Eppler & Mengis, 2004). The Internet is packed with useless webpages and documents. Therefore, the document ranking algorithms play an extremely important role for filtering documents in an information retrieval system.

Brin and Page (2012) proposed PageRank (one of the most well-known webpage ranking algorithms) which borrowed the idea from citation analysis. PageRank states that the importance of a webpage is estimated based on the importance of the entirety of webpages which contain hyperlinks referring to it. The more important the referrers are, the more significant the referral is. It is a good idea to determine the relative significance of a website by considering its referrers. It's easy to know the significance of an academic journal since there is already an established mutual recognition system in academia. For instance, the rank of a journal from the Web of Science database is known. However, for the Internet community, there was no such system back in 1998. Therefore, in order to get the PageRank value for the entire Internet, the value for a list of websites must be calculated. The question was "where was the start point?" If two websites, A and B, refer to each other, the importance of A is calculated based on the importance of B, and the importance of B is computed by the importance of A. This is an endless loop. Langville and Meyer (2011) articulated how Brin and Page resolved this problem in their book *Google's PageRank and Beyond: The Science of Search Engine Rankings*. The founders of Google considered each webpage on the Internet to be a state and the transitions among webpages with hyperlinks to be state transitions. The probability from webpage A to webpage B (A has a hyperlink referring to B) was calculated as the sum of all the hyperlinks in webpage A divided by 1. Therefore, the vector corresponding to webpage A was obtained. With the adjustment, the transition matrix was the sum of all the webpages' vectors on the Internet. Given a random vector with the dimension of the number of total webpages on the Internet, multiple it with the transition matrix several times until the vector became stabilized, then the value of the vector is the PageRank value for the corresponding webpage. The PageRank algorithm played an extremely significant role in online information retrieval and helped lay the foundation of Google.

Back to academia, Daniłowicz and Baliński (2001) viewed each document as a state of a Markov chain, and the ranking of the document was calculated based on the transition probabilities. To be more specific, the transition probability from one document to another was calculated by using their Cosine similarity. The ranking of the documents was determined by the steady-state probability distribution which indicated the first rank as the document presenting the set best and the last rank document as the worst representative of the set. The degree of importance for a document equaled its degree of representative.

2.2.3 Markov chain-related techniques and their variations

Apart from the traditional Markov chain method, many Markov chain-related alternatives have been proposed to deal with the dynamic and ever-changing environment of the information retrieval field.

2.2.2.1. Hidden Markov chains and machine learning

An important variation is the hidden Markov model (HMM). In a real research study, the states of a Markov chain are not always observable. However, the outputs from the unobserved states can be obtained. Each state has a probability distribution of outputs.

Machine learning techniques have always been used in conjunction with the hidden Markov chains. Generally speaking, it was difficult or even impossible to accurately estimate the hidden states which only rely on the outcomes. Consequently, ideas brought from machine learning helped to solve the problem. To get better predictions, a training data set was required to estimate the parameters in the system. In the following section, this paper will discuss research that has been applied to the hidden Markov chain and the machine learning techniques at the same time.

In the field of information retrieval, Miller, Leek, and Schwartz (1999) proposed an HMM-based information retrieval system. In the system, the observed outputs were the queries created by the users, while relevant documents were the hidden states. Given a query, the probability for each document in the document set to be related to the query could be estimated. The authors tested the proposed model on TREC-6 and TREC-7. The results showed that the suggested model outperformed the *tf***idf* based models in terms of matching query and document.

Document classification is a hot topic. Yi and Beheshti (2008) applied the HMM in text classification. The authors used the Medical Subject Headings, controlled vocabulary for medical journals and books, as the training dataset. The sources of information were viewed as the states of a Markov chain, and the state transition denoted texts transited from one source to another. The results showed that the performance of the algorithm based on the Markov chain was comparable to other machine-learning models such as support vector machines, Bayesian belief networks, and artificial neural networks. Frasconi, Soda, and Vullo (2002) approached this problem in finer granularity. They pointed out that books and magazines were documents with multi-pages and that the structural information could not be ignored as it was in the news documents. It was not appropriate to represent a multi-page document as a vector which regarded a document as a “bag of words.” As a result, an HMM was introduced to deal with the structural information of documents. It regarded each page as a hidden state and the object was to label page structural information. Finally, the Viterbi algorithm was used to categorize the pages with a structural information label.

Information extraction refers to “any process which selectively structures and combines data which are found, explicitly stated or implied, in one or more texts” (Cowie & Lehnert, 1996). In an information retrieval system, adopting information extraction techniques helps the user get more relevant information than using an entire document. Seymore, McCallum, and Rosenfeld (1999) designed a model to extract information from the header of computer science research papers. The header contains all the information prior to the main body, includes the title, author names, and affiliations. The HMM was used to predict to which field each word extracted from a header belonged. The results specified that the proposed model achieved an accuracy of 92.9% in general.

In order to achieve better predictive power, a large and comprehensive labeled training set is necessary for estimating the hidden states accurately. However, in reality, it’s too impractical to obtain a perfect training set. Many researchers have tried to build a simpler model with few states for alleviating the effect of insufficient training data. The problem was, that although it made the parameter estimation more accurate, the prediction rate was not improved since the amended model does not present the data adequately (Freitag & McCallum, 1999). It was a dilemma of building a model with sufficient states but inaccurate parameters and a model with precise parameters but poorly representing the data set. Freitag and McCallum (1999) used a statistical technique called “shrinkage” to resolve the problem. Shrinkage helped to shrink the parameter estimations from a complex model to a simple model. The results demonstrated that the performance of an HMM-based information extraction model improved significantly by adopting shrinkage when there was limited training data.

In a word, the HMM provided a solution to the problem when the states were invisible. In information science, the outputs were always terms, and the states were documents or document segments. Other applications included handwriting recognition (Chen et al., 1994; Mohammed & Gader, 1996), speech recognition (Rabiner, 1989), data compression (Cormack & Horspool, 1987), and so on.

2.2.2.2. Higher-order Markov chains and information visualization

One of the primary purposes of utilizing the Markov chain method is to predict the probability of future state distribution. The accuracy of prediction is limited since they only rely on the information of the current situation (Deshpande & Karypis, 2004). In order to achieve better predicting performance, Markov chains with higher-order have been proposed to consider more preceding information (Borges & Levene, 2005). For example, a second-order Markov chain predicts the future state based on the past two states.

Xie and Joo (2005) presented a fifth-order Markov chain in modeling the most frequently applied search strategies. However, a higher-order Markov model dramatically increased the computational cost since the number of states increased significantly (Deshpande & Karypis, 2004). In Xie and Joo's case, despite the fact that there were only thirty-one participants and sixty search tasks it was considered acceptable for this experiment. For a large data set that comes from a Web portal or search engine, it was impractical to use a high order for better prediction without considering the burden of computational capabilities.

Many attempts have been made to design a compromised model with acceptable computational cost while retaining decent prediction accuracy (Deshpande & Karypis, 2004). Support-pruned Markov model removed the states which got low support from the training data set. The Confidence-pruned Markov model retained only the most frequently taken actions. And the error-pruned Markov model kept the states which had the lowest error rate during the validation process. The authors then conducted a comprehensive experiment for comparing these models with the first-order model, the second-order model, the third-order model, and all K^{th} model which combined the states from several different higher-order models. The results showed that the proposed error-pruned Markov model had the highest prediction accuracy while it had an acceptable computational cost. Borges and Levene (2005) argued that information retrieval was a dynamic and ever-changing process. There was no one Markov chain model able to represent the users' navigation behavior in all different situations. In view of that, they examined the ability of different higher-order Markov chain models in three data sets. The results suggested that some websites need a long history to model users' behavior while others only require a short history.

Instead of selecting the Markov model with the proper order, Pitkow and Pirolli (1999) defined the states in a Markov chain themselves. The defined states were referred to as the longest repeating subsequences. It suggested that the users' navigation paths that occurred repeatedly should be noticed. The results indicated that the model based on the longest repeating subsequences was able to make accurate predictions. Incorporating the longest repeating subsequences in other models would increase the predictive capabilities.

2.3. Transaction log analysis and data mining

Web transaction logs are used to record the interactions between users and Web-based systems in the database. An entry of the transaction log records an action that includes an Internet Protocol (IP) address, a user machine cookie, a timestamp indicating the action time, a referrer directing to the current page, a referral which is the present page, and a session identifier. Since the transaction log records almost every user's interaction, it has always been used for studying the users' behavior and improving or reorganizing the Web-based systems. This section focuses on summarizing the studies on the transaction log of Web-based systems.

2.3.1. Transaction log analysis in information retrieval systems

Information retrieval (IR) systems are designed to meet users' information needs by providing anything that might satisfy them, such as documents, webpages, audios, videos, images, and so on. Many IR models have been proposed, such as Boolean models, vector space models, and probability models. IR systems can also be classified by the users or usage, including online public access catalogs (OPAC), digital libraries, search engines, and so on.

Chowdhury (2010) defined OPAC systems as systems that "allow users to search by using typical bibliographic keys such as author names, titles of documents, subject descriptors and keywords, and at the end of a search they produce a list of documents, with some bibliographic information and a call number." The OPAC offers high quality and peer-reviewed information sources with a keyword search feature. As an extension of the traditional library, Witten, Bainbridge, and Nichols (2009) defined the digital library as a library with a digital collection, such as text, videos, audio, and other information stored in electronic formats.

The most robust IR systems, search engines, are comprehensive information systems that help users to find various kinds of information. Search engines can be further classified into two categories, query-based search engines such as Google and subject directories such as Yahoo.

The interactions between users and IR systems have been widely studied for both understanding users' information needs and improving systems' performance. The transaction logs provide an unobtrusive way to study the interactions. In this paper, the analysis of interactions between users and IR systems is discussed. Jansen (2006) suggested that the analysis of transaction log data should be conducted at the term, query, and session-level; the following section is organized applying these three levels.

2.3.1.1. Term level

Jansen and Pooch (2001) defined *term* as “a string of characters separated by some delimiter such as a space, a colon, or a period”. The analysis on the term level focused on the descriptive statistics.

Spink, Wolfram, Jansen, and Saracevic (2001) analyzed the distribution of term frequency in a Web search engine. They found that 57.1% of terms were used only once, 14.5% were used twice, and 6.7% were used three times. Their analysis indicated that only a small fraction of terms were frequently used in the search, whereas a large number of terms were seldom used. The distribution of term frequency fitted a classic informetric law, Lotka’s law. Taking the logarithm of both the term frequency and the term rank, a Zipf –like distribution was identified.

Wolfram (2008) conducted a comparison study on four different IR systems: a bibliographic databank, an online OPAC system, a search engine, and a health search system. He noted that all four IR systems include heavy usage of articles, prepositions, and conjunctions. The terms submitted to the bibliographic databank and OPAC were more academic-related terms, although the OPAC is primarily designed to manage monographs. The search engines tended to deal with the terms related to multimedia, software, education, and adult-oriented topics. Obviously, in the health system terms used reflected the popular health topics like cancer. In addition, the term frequency distributions of four IR systems could also be visually identified as a Lotka’s pattern.

Moreover, Zhang et al. (2008) analyzed the frequently used health-subject terms and their co-occurring query terms. They found that the query terms which occurred with the subject term in the same query usually could be clustered into two to four clusters with large concentrated groups of terms and several specialized clusters. This suggested that the co-occurred terms were usually the terms of medicines/drugs which were usually used in the disease related to the subject terms, the terms related to symptoms, and the terms related to the causes. They also mentioned that there was a gap between the terms adopted by users in describing information needs and the formal medical terminology used by the professionals.

2.3.1.2. Query level

In IR systems, the users directly selected queries to represent their information needs. If the returned results failed to meet the information needs, they tended to modify the queries until they obtained the needed information. Therefore, the query analysis might be the most effective and efficient approach to revealing the users' information search behaviors (Wolfram, 2008).

"A query is usually a string of zero of more characters entered into the Web IR system" (Jansen & Pooch, 2001). They also indicated the early analysis at the query level concentrated on discovering the length of the query, the complexity of the query, and the failure rate. More specifically, query length was the number of terms used in a query. Query complexity observed the use of advanced search features including the use of Boolean operators and the use of term modifiers. Failure rate measured the rate of generating improper queries which can't be processed by the system. These three measurements represented the proficiency of users in different dimensions.

There are some contradictions among the findings in terms of the complexity of the queries. Wolfram (2008) noticed that users from OPAC tended to submit a more complicated query with more terms than others. This finding agrees with our knowledge since the OPAC users are more professional and have a better knowledge of using operators. In addition, Wolfram found that users of a search engine tended to read more pages per query. However, Jansen and Spink (2006) argued that most users of search engines only read the first page. This argument was also supported by the findings of a previous study conducted by Spink, Jansen, Wolfram, and Saracevic (2002).

Besides text-based information retrieval, multimedia information retrieval plays a significant role in the age of the Internet. Ozmutlu, Spink, and Ozmutlu (2003) investigated the users' multimedia searching behaviors in Excite. From the transaction logs ranging from 1997 to 2001, they noticed that compared with the queries for text-based information, the average length of multimedia-related queries were much longer (containing almost three times more terms). Other characteristics of multimedia searching include: people were more willing to search audio-related information than video and image information; and users tended to use fewer but more complicated queries in a session which suggested the users were becoming more proficient in searching as time went on. However, a more recent study did not coincide with some of these findings. Tjondronegoro, Spink, and Jansen (2009) examined the search behaviors of Dogpile users in 2006. They found the image search was the most prominent search within the multimedia searches, followed by the audio search and the video search. Compared with the study in 2003, the user's query was shorter (2.6 terms per query in 2006 while 4.3 terms per query in 2003) and the topics of search changed from large entertainment to broader topics such as medical, sports, and technology.

Query classification facilitated the studies on transaction logs in several ways. On the one hand, it helped researchers better understand users' information needs and information behaviors.

On the other hand, it shed light on session boundary detection (Jansen, Spink, Blakely, & Koshman, 2007). Spink, Hansen, and Ozmultu (2000) defined four mutually exclusive classes for queries classification in the search engine Excite. The unique query was the query submitted by a user to represent his information need in the first place. If the results retrieved from the unique query failed to satisfy the information need, the queries modified by the user based on the unique query were called modified queries. Other queries were classified as the next page, which stood for activities such as reading the results from the next page after submitting a query, and relevance feedback, a function provided by Excite for finding the relevant link. On top of that, He, Göker, and Harper (2002) defined a more complicated query classification with at least eight mutually exclusive classes: browsing, generalization, specialization, reformulation, repetition, new, relevance feedback, and others. The browsing corresponded to the next page category from Spink, Hansen, and Ozmultu's (2000) classification, and the new query corresponded to the unique query. However, He, Göker, and Harper further classified the modified query into three sub-classes: generalization, specialization, and reformulation. The generalization query stood for the activities of seeking more general information on the same topic, while the specialization query referred to looking for more detailed information about the topic. The reformulation query was used to modify the original query for getting the information about the same topic.

In addition to the query classification schemas proposed according to the search patterns, queries were also classified based on the topic. Spink et al. (2001) proposed a scheme with several categories: entertainment and recreation; sex, pornography, and preferences; commerce, travel, employment and economy; society, culture, ethnicity, and religion; education and the humanities; the performing and fine arts; government; and unknown and incomprehensible. To cope with the mutability nature of topics, Jansen, Booth, and Spink (2008) put forward a scheme for classifying queries in terms of the user's goals derived from the work in Rose and Levinson (2004). The categories included informational searching, which described the queries submitted for searching content related to the topic; navigational searching, which was defined as the queries proposed to locate a particular website; and transactional searching, which referred to the searching activities for finding a website to attain products. They discovered that the terms submitted for informational searching accounted for more than 80% of the total queries, while the other two categories constituted 10% for each.

2.3.1.3. Session level

The session refers to "the entire sequence of queries entered by a searcher" (Jansen & Pooch, 2001). Since sessions could be understood as users, the analysis on the session level is equal to the users' analysis.

Wolfram et al. (2009) examined the sessions from three different IR systems: academic website, public search engine, and consumer health information portal. By conducting a clustering analysis, three unique session behaviors were identified across all three IR environments. Most sessions could be defined as "hit & run", which represented users who used uncommon terms and didn't modify the queries after the first attempt. This type of user could be efficient users who met their information needs quickly or inefficient users who gave up immediately after failure. Other sessions could be classified as "popular & focused" and "long & varied". "Popular & focused" corresponded to the users who used common queries and conducted medium-length sessions. "Long & varied" characterized the users who performed the longest sessions and were more willing to modify the queries than the other two groups. Han et al. (2015) also obtained three clusters from an image-based IR system. The sessions of the first cluster constituted 97% of the dataset. It represented the users who performed a quick search and viewed few results. The second cluster included 2.52% of the sessions and represented sessions with medium values of session characteristics. The last cluster only accounted for 0.11% of the sessions and denoted the proficient users.

In addition to clustering the sessions, the actions recorded in a session were regarded as sequential search actions conducted by the user. Han et al. (2015) identified several usage patterns. In the system, the average number of actions performed by a user was 21.4, and the three most frequently occurring actions related to browsing and simple searches (single item view, page of results request, and simple search) accounted for more than 86% of the total actions. It suggested that most users only conducted simple search functions and were reluctant to apply advanced search functions such as facet search during the information seeking.

A major limitation of transaction log analysis is that logs cannot record users' real perceptions during the search (Kurth, 1993). Kurth stated that the transaction log "cannot measure the information needs that users are unable to express in the search statements that they enter into online systems." To solve the problem, several qualitative data collection methods have been incorporated in conjunction with log analysis. Xie and Joo (2010) integrated qualitative data from the prequestionnaire, information interaction diary, think-aloud protocol, and postquestionnaire as complements of the transaction log data.

2.3.2. Challenges in Transaction Log Analysis

The transaction log analysis has proven to be efficient and effective in studying the interactions between users and IR systems. However, there remain many challenges associated with analyzing log data from public servers. This paper considers three major challenges, including the problem of detecting boundaries between sessions; the dilemma of preserving users' privacy and maintaining the utility of data; the enormous size of transaction logs associated with the burden for both human process and analysis, and computational requirements.

2.3.2.1. Session identification

The most prevalent challenge in the studies on transaction logs is how to define the boundary of the sessions. The session is the fundamental unit to reflect a user's information need. Abundant analyses were conducted at the session-level such as the studies on query reformulation, topic classification, and so on. As a result, it is really important to identify the sessions before data analysis. Some traditional IR systems such as OPACs and bibliographic databases provide a session identifier, which distinguishes the queries submitted by a user from the entry point to the exit point. The function facilitates researchers in session analysis. However, many IR systems only identify users by relying on the Internet Protocol (IP) address and machine cookies.

Various attempts have been made by the researchers in order to solve this problem. Wolfram (2008) summarized two categories of session boundary detection methods. The first category was based on content analysis, and initially analyzed the topics of queries. If the consecutive queries submitted from the same IP address share the same topic, the queries are determined to belong to a session. Another method for identifying sessions depended on the timing characteristics of the database. It used temporal cut-off points to lay the boundaries between sessions. The weakness of the first method was that it assumed a session only covered one topic without considering the possibility that users may change search topics in or during one search (Spink et al., 2006), whereas the second method only took into account the temporal characteristics of the dataset but failed to consider other characteristics.

Many studies defined a general cut-off point. Göker and He (2000) argued that an optimal session interval was able to ensure that most of the sessions contain a reasonable number of activities. They decided that sessions with one to six activities could be considered as reasonable in the Reuters log. By observing the distribution of sessions with different session intervals, they found that the sessions with six or fewer activities constitute the majority (81%) of sessions when the session interval is within ten to fifteen minutes. Combining the results from the experiment of human identification, the authors concluded an optimal session interval should be between eleven to fifteen minutes. The results from a similar experiment conducted on the Excite logs supported their suggestions in defining the session cut-off point (He & Göker, 2000).

However, Wolfram (2008) pointed out that "an arbitrary assignment of a cut-off point based on past studies without considering the characteristics of the datasets themselves would have been shortsighted." The author made two assumptions about the logs he examined. First, a session should not cover several days. Second, the cut-off point for a dataset should not only consider the suggestions proposed by the earlier studies but also take into account the characteristics of the currently examined dataset. He further argued that the cut-off points should be less than the mean value of session lengths based on a finding discovered by Spink and Jansen (2004). Spink and Jansen learned that although the average session length is two hours, most of the sessions were less than fifteen minutes from a system with the session identifier. As a result, Wolfram proposed a session boundary detection method where the sessions should be separated at the point of 80% of the average session length. In the study, the cut-off time for the Excite search engine was 1074 seconds and contained 1.8 queries per session; and for a health-subject IR, system was 229 seconds and included about 2.3 queries per session. Beyond that, Han and Wolfram (2015) incorporated Kernel regression to smooth the distribution of time intervals between actions. The optimal cut-off point was 1585 seconds for an image-based digital library. Murray, Lin, and Chowdhury (2006) introduced an approach in terms of the burstiness of users' activity. They used a variant of the hierarchical agglomerative clustering algorithm in cutting sessions with a usual time interval.

In respect to the content analysis-based session identification method, Jansen, Spink, Blakely, and Koshman (2007) compared three-session identification methods, using IP address and cookie; using IP address, cookie and a temporal limit on intrasession interactions; and IP address, cookie and query contents. In the last method, the authors applied a query classification method. They first classified queries into six mutually exclusive groups: assistance, content change, generalization, new, reformulation and specialization. Consequently, the query with a unique IP address and the cookie was defined as a new session. The query with the same IP address and cookie but had no common term with the previous query was identified as a new session as well. Comparing the results from the proposed method with the results from human identification, they concluded the method incorporating content analysis yielded 95% accuracy in terms of session boundary detection.

In addition to the session segmentation algorithms which solely rely on each proposed paradigm, He, Göker and Harper (2002) invented an automatic session identification method that incorporated the content analysis and the temporal cut-off points based on the Dempster-Shafer theory. The Dempster-Shafer theory has been widely accepted by the information processing community for combining different evidence. Compared with other session boundary detecting methods, including those which solely rely on the temporal analysis as well as those which consider the content analysis, the authors concluded that the proposed method significantly outperformed the other two methods in terms of segmenting sessions.

In a summary, although there is no universal and perfectly accurate algorithm for the session identification, there is always an optimal way to segment the sessions: considering the temporal characteristics of the dataset; considering the content analysis in terms of topic change; or incorporating the two strategies at the same time.

2.3.2.2. Privacy preservation

Another critical challenge along with transaction log analysis is users' privacy preservation. In the era of the Internet, protecting the privacy of users has become a significant issue for studies including personal information (Zimmer, 2013). The queries submitted by users can reveal the interests, preferences, search strategies, and many other characteristics of the users (Cooper, 2008). Jones, Kumar, Pang, and Tomkins (2007) further pointed out a threatening fact that they are able to identify a set of users with a decent probability by simply using the query data.

The challenges of privacy preservation for the transaction log studies have attracted the attention from researchers since 1993 (Kurth, 1993). Attempts, including removing IP address and machine cookies, ventured to make the users anonymous (Jones, Cunningham, McNab, & Boddie, 2009). Cooper (2008) tried to combine the technical and policy methods to reduce the likelihood of personal information leaking significantly while preserving the utility of data. The author reviewed several privacy preservation techniques. Log deletion deleted the users' query logs such as queries, user identifiers, timestamps, and all other information from the database. Although this method fully protected the users' privacy, there was no way to utilize the data by researchers or industries. Hashing queries converted the string data to a hash value that was difficult or even impossible to restore to the original value. These methods helped protect privacy because of the high difficulty of restoring the original query. However, difficult was not equal to absolutely impossible. In some cases, if one obtained the previously released query data, it was possible to convert the hash values to the original queries. Identifier deletion removed the IP addresses and machine cookies from the transaction log. The problem of this was that by removing the identifier it made it impossible for the systems to provide any function that needed user profiles, such as a personalized hub or recommendation. It also caused a problem related to a subpoena. If courts or government authorities issue a subpoena, it was hard to locate the subject since identifiers had been deleted. Hashing identifiers referred to a method for converting all external identifiers of a query to the hash values. Nonetheless, there remained the possibility of personal information leaking. The queries submitted by a hash identifier could still be linked with the possible identifiable information within the content of queries. Other techniques involved: scrubbing query content which removed identifiable information such as phone number, name, Social Security numbers and so on in the queries; deleting infrequent queries which might have been used to identify users; and shortening the session which reduced the possibility of revealing identity. The author then analyzed the policies related to privacy reservations such as privacy law, privacy policies, confidentiality and licensing agreements, consent, and Institutional Review Boards (IRB). A more recent attempt was made by Navarro-Arribas, Torra, Erola, and Castellà-Roca (2012). The authors applied a microaggregation method to the anonymization of transaction logs. They concluded the proposed method helps the *k*-anonymity of the users in the log while preserving the utility of the data.

2.3.2.3. Sampling

A narrowly recognized challenge associated with the studies on transaction logs is related to the sampling process. Due to the nature of the transaction log from Web-based IR systems, there is always an enormous amount of data to be analyzed. For example, a transaction log extracted from an image-based digital library contained more than twelve million actions within three months (Han et al., 2015). Several data cleaning methods were performed before data analysis, which included excluding overlapping actions with the same IP address, eliminating sessions with an unreasonable number of actions (more than 1000 actions), and system-generated actions. However, there were still a significant number of records (325,059) to be analyzed. Things are getting worse in search engines. A transaction log analysis of AltaVista conducted by Silverstein, Marais, Henzinger, and Moricz (1999) constituted approximately one billion actions!

Regarding that, Ozmutlu, Spink, and Ozmutlu (2002) indicated that there was a need to find sampling techniques to reduce the transaction log dataset to a manageable size while retaining the statistically representative characteristics of the set. The authors compared two sampling strategies, systematic sampling, and Poisson sampling. Systematic sampling selected data points according to a fixed periodic interval without considering the patterns under the data. It worked well with the query data since there were no patterns in the queries. However, for tasks such as selecting the session data and the data used to reveal the search patterns, systematic sampling was not competent to accomplish the job. The authors applied the Poisson sampling strategy, which selected data points by ignoring a random number of observations that have a Poisson distribution to transaction logs. They concluded that, compared with systematic sampling, the Poisson sampling was statistically efficient to represent the characteristics of the transaction log data.

Another potential technology that could be applied to the sampling process is temporal analysis. The temporal analysis was defined as “the analysis of time-ordered sequences of unit vectors” (Fisher, Lewis, & Embleton, 1987). It examined the pattern variations from the time dimension. If one demonstrated that the patterns or characteristics of the transaction log data to be tested in the study remained stable in terms of temporal properties, it was practical to select a segment of data, such as the queries submitted from one day log of a six-month log for conducting a study on the particular patterns or characteristics. Beitzel, Jensen, Chowdhury, Frieder, and Grossman (2007) examined query topic patterns on a daily, weekly, and monthly basis. They concluded that although some topics like music, movies, and entertainment showed a dramatic change over time, other topics including personal finance and porn remain stable during a certain time. These findings shed light on the possible application of temporal analysis-based sampling strategies to the studies of the transaction log. For future studies, there is a need to discover the temporal characteristics of other patterns such as search patterns and traversal patterns in terms of facilitating the sampling strategies for the transaction log studies.

2.3.3. Techniques in Transaction Log Analysis

The data extracted from transaction logs are a gold mine. Transaction logs can be used to extract the characteristics from both users' side and systems' side. Many techniques have proven successful in analyzing the transaction log. In this paper, three major techniques associated with the transaction log analysis are discussed, including data and text mining, network analysis, and informetric.

2.3.3.1. Data and text mining

The primary purpose of data mining is to discover the models from data (Leskovec, Rajaraman, & Rajaraman, 2014). Data mining is defined as "the process that uses statistical, mathematical, artificial intelligence and machine-learning techniques to extract and identify useful information and subsequently gain knowledge from large databases" (Turban, Aronson, Liang, & Sharda, 2007). Obviously, the transaction log data is large enough data to be mined. In this paper, two data mining techniques including clustering and pattern mining are discussed.

Clustering is defined as “the process of examining a collection of points, and grouping the points into ‘cluster’ according to some distance measure” (Leskovec, Rajaraman, &Ullman, 2014). The object is to classify the points with a small distance into the same cluster while keeping points in other clusters with a large distance. The major difference between classification and clustering is that a classification system requires a set of predefined categories while clustering doesn’t necessarily need it. The techniques of clustering have been widely adopted for grouping a huge amount of data, and can be applied in grouping transaction log data.

Kim and Seo (2006) designed a FAQ retrieval system. In the system, the queries were clustered into the predefined FAQ categories. Then matching the newly submitted query with each cluster, Beeferman and Berger (2000) applied an agglomerative clustering algorithm to cluster queries and the URLs which the user viewed from the returned list. The queries and URLs clustered together were not based on the content but the co-occurrence within one session.

The limitation of clustering methods is that although a clustering algorithm provides a way to categorize items, it does not deliver the information about the relationships between clusters, nor the relationships between terms which belong to different clusters (Zhang et al., 2008). As a result, Zhang et al. introduced an information visualization technique, the multidimensional-scaling approach (MDS), in order to unravel the problem. The MDS was able to project the semantic relationships of the terms from a high-dimensional space into a low-dimensional space (usually three-dimensional). With the employment of the MDS, the intensity of semantic connections between terms and clusters could be determined by the distance between them.

In addition to the term clustering, the sessions (herein, sessions are equal to users) can also be clustered. Wolfram, Wang, and Zhang (2009) clustered sessions from three different IR systems. The sessions were clustered in terms of ten characteristics: session length, the average number of terms used per query, average term popularity, average query interval, average term use frequency, the average number of pages viewed per query, number of searches using Boolean operators, the average number of unrecognized or nonstandard words, average number of stopwords, and average term number changes. Han and Wolfram (2015) used the K-means algorithm for session clustering. Fourteen session characteristics were identified, which included: session actions, session length time, queries per session, average terms per query, image page visits, landing page visits, page result requests, the average time between actions, the average time between queries, the proportion of browsing actions, average image pages per query, print views, the proportion of landing page visits, and proportion of page result requests. To reduce the characteristics with high correlations, exploratory factor analysis was conducted and six characteristics with high loading were retained. How to define the characteristics of the session was a challenge. In the studies discussed above, the characteristics were identified in two ways, directly observed characteristics and characteristics derived from the data.

Pattern mining, also called frequent pattern mining, was first proposed by Agrawal, Imieliński, and Swami (1993) for analyzing customers' buying behaviors using a large set of transaction data. Han et al. (2007) defined the patterns as "item sets, subsequences, or substructures that appear in a data set with a frequency no less than a user-specified threshold". They further noted that a frequent itemset was a set of items that appeared repeatedly together in the dataset. A frequent subsequence was a sequence of actions occurring frequently in the dataset. A substructure referred to "different structural forms, such as subgraphs, subtrees, or sublattices, which may be combined with item sets or subsequences."

From the perspectives of the studies on the transaction log of IR systems, the pattern mining analysis has always been conducted in three facets: analysis of subsequences, item sets and substructures. The frequent subsequences analysis refers to the studies on sequence patterns in the Web transaction log studies (Yang et al., 2001; Han & Wolfram, 2015). The frequent item sets analysis stands for the frequently occurred queries during the search (Fonseca et al., 2003; Fonseca et al., 2005). And the substructures analysis signifies the studies on the users' traversal patterns and the structures of the IR systems (Srikant & Yang, 2001).

Since the search activities represent a series of sequential activities that took place in the IR systems during information seeking, sequence mining has been largely accepted by the researchers in modeling the users' traversal activities. Han and Wolfram (2015) extracted the sequence patterns from an image-based digital library using the CM-CLaSP algorithm. They found that the most frequent pattern was the combination of simple search followed by the individual item views. It indicated that most of the users tended to search simply: search the needed information and view the items from the returning list. Rieh and Xie (2001) examined the query reformulation patterns in a Web search engine. Six patterns were identified: specified reformulation in which users attempted to specify their original submitted queries; parallel reformulation in which users attempted to submit another query which shared common characteristics; dynamic reformulation referred to a complicated sequence of actions that users tried to specify or generalize their query; alternative reformulation in which users tried an alternative query to fulfill their information needs; and format reformulation in which users tried to format their queries to meet systems' query requirements. Jansen and McNeese (2005) identified 171 unique patterns and further isolated ten patterns with searching assistance. Zhang et al. (2004) identified three searching patterns (focused searching, exploratory searching, and unsystematic searching) in a medical IR system. Zhou, Hui, and Fong (2006) tried to improve the performance of a famous pattern mining algorithm, Web Access Pattern tree, by reducing the computational burdens. The proposed method, Conditional Sequence Base mining algorithm, was proven effective in pattern mining.

The extracted sequence patterns can be used to enhance the performance of IR systems. Yang et al. (2001) identified frequent website access patterns and designed a website storage model to fetch and cache the websites which may be read by users. Sarukkai (2000) incorporated pattern mining and a Markov chain model for link prediction. The recognized patterns were modeled by the Markov chain and the probability of transition between links was estimated.

Sequence mining is usually used in digging for patterns of a series of ordered information-seeking activities. An alternative pattern mining method, association rule learning, only considers the co-occurrence activities, and has been proposed to deal with the problem of mining unordered activities. Query recommendation is an important function of search engines. It helps users to clarify their information needs or to reformulate their queries by making suggestions. The queries from existing transaction logs, which reflect users' information needs directly without the need to consider their sequence, well suits the query recommendation studies. Many query recommendation and query expansion studies have incorporated the association rule (Fonseca et al., 2003; Fonseca et al., 2005). Shi and Yang (2007) implemented an improved association rule learning method in query recommendation. They compared the relatedness between the submitted query and existing queries extracted from a transaction log by the association rule learning method. Then they returned a list of the most related queries to the user. The results signified that the proposed method outperformed the temporal correlation model in terms of query recommendation.

Many pattern mining methods have been successfully applied to discover the users' traversal patterns in IR systems. However, a limitation of pattern mining has been identified, that is, no pattern can consistently represent the users' traversal behavior. As the transaction log keeps expanding, the users' searching behavior may change as well. Furthermore, the system structure may be changed due to system updating (Lee & Yen, 2008). The patterns extracted from previous transaction logs may be out of date. Lee and Yen (2008) introduced two incremental data mining algorithms, the *IncWTP* algorithm for coping with traversal pattern changes, and the *WssWTP* algorithm for managing system updates.

2.3.3.2. Network analysis

Network analysis is a tool borrowed from social networks. In social science, the network analysis "provides an answer to a question that has preoccupied social philosophy since the time of Plato, namely, the problem of social order: how autonomous individuals can combine to create enduring, functioning societies" (Borgatti, Mehra, Brass, & Brass, 2009). In the analysis of transaction logs, network analysis can be used to visualize the characteristics and patterns of the log data.

Han and Wolfram (2015) examined the relationships between two adjacent search actions and the network constructed through the relationships in an image-based digital library. The actions were nodes, and the co-occurrence of actions was defined as a relationship in the network. The resultant network showed that the most important actions (with the highest betweenness centrality value) were using simple searches and viewing individual items. Benevenuto et al. (2009) constructed a network based on friendship relationships extracted from transaction logs to represent the topology of Orkut but did not further analyze it. Saha Roy et al. (2016) conducted a detailed complex network analysis of query terms which shed light on the transaction log analysis. The network was built on two kinds of term co-occurrence in the query: local co-occurrence, which meant only neighbor terms were considered as important; and global co-occurrence, which referred to a relationship that an edge between terms would be added if two terms occurred in the same query regardless of their positions. Many measurements from network analysis were also used to exhibit the characteristics of the query term data.

In summary, network analysis strengthened the study on transaction logs in multiple ways. It identified the influential nodes in the network. These nodes could be the search actions (Han et al., 2015), query terms (Saha Roy et al., 2016), or users on the website (Benevenuto et al., 2009). It represented the relationships between nodes in terms of centrality. The relationships were the frequently conducted search action pairs in a session (Han et al., 2015) and the frequently used query terms (Saha Roy et al., 2016). The network analysis also shed light on the sampling process. A query terms network remained stable when there were at least 100,000 terms in terms of the degree distribution. The implications might be that the characteristics or patterns of the query logs would remain stable when the sample size was large enough. Finally, it visually represented the characteristics and patterns of log data.

2.3.3.3. Informetric

Informetric was defined as “the study of the quantitative aspects of information in any form, not just records or bibliographies, and in any group, not just scientists” (Tague-Sutcliffe, 1992). Another similar definition provided by Wilson (1999) was the method for “dealing with the measurement of information phenomena and the application of mathematical methods to the discipline’s problems.” In the informetric field, there are three most famous laws to mathematically represent the different aspects of information. These are Lotka’s law, Zipf’s law, and Bradford’s law. Many studies have been conducted to examine whether the patterns discovered from transaction logs were able to be mathematically modeled. The implication of the studies was that a fitted model representing the transaction log data can yield simulations of user behaviors (Ajiferuke & Wolfram, 2004).

Ajiferuke and Wolfram (2004) conducted a comprehensive study of informetric modeling query data. The data were used to represent query and search behaviors (term frequency, terms used per query, pages viewed per query, query frequency, queries per session, and page viewed per session). They concluded that a classic Zipf model was not adequate to model the transaction log data since a huge dataset might cause data distribution to be highly skewed, whereas it demonstrated that the generalized inverse Gaussian-Poisson distribution fit the data best in terms of the goodness-of-fit tests. However, Benevenuto et al. (2009) discovered the distribution of session lengths from a social network could be fitted in a Zipf model. Ajiferuke, Wolfram, and Famoye (2006) also examined the impacts of sample size on informetric modeling. It was widely recognized that the size of the dataset could influence the outcomes from goodness-of-fit tests since the distribution of data was getting more skewed as the data size increased. However, the distribution characteristics of the averages of terms used per query and session remained reasonably the same as size increased. In contradiction, Saha Roy (2016) observed that the degree distribution of terms in a network could consistently fit in a two-periphery distribution when there were more than 100, 000 terms.

Besides the query data, visitation data and resource usage data extracted from transaction logs could also be modeled by using informetric methods. Ajiferuke, Wolfram, and Xie (2004) concluded that the distribution of the number of visitations from a specific IP address could be reflected by a Zipf model, while the distribution characteristics of the number of requests for a specific resource could be represented by a generalized logarithmic series model. Davis (2004) analyzed the referral links data (a webpage refers a user to the targeted webpage) from the American Chemical Society servers. Although most referrals originated from several academic-related sources (OPAC, library e-journal list, and bibliographic databases), a considerable number of referrals came from a lot of different sources. The author found that the distribution of types of referrals followed Lotka's law (inverse-square law).

Even though the transaction log data have been proven to possess the characteristics that could be modeled by informetric laws it is still difficult to model characteristics of transaction logs other than the descriptive characteristics like the average terms per query (Wolfram, 2008).

2.4 Online health information needs and seeking

According to a national survey conducted by the Pew Research Center's Internet & American Life Project (http://www.pewinternet.org) in 2014, 35% of adults in the U.S. had searched on the Internet to find out what disease or health condition they or someone else might have. Among them, 35% did not further consult a professional. It proved that the Internet was an important source for health information consumers and was able to affect consumers' decisions. Therefore, it is vital to understand how consumers seek health information online.

2.4.1 Online Consumer health information-seeking behaviors

Health information seeking can be defined as:

> Any non-routine media use of interpersonal conversation about a specific health topic and thus includes behaviors such as viewing a special program about a health-related treatment, using a search engine to find information about a particular health topic on the

> Internet, and/or posing specific health-related questions to a friend, family member, or medical practitioner outside the normal flow of conversation. (Niederdeppe, Hornik, Kelly, Frosch, Romantan, Stevens, Barg, Weiner, & Schwartz, 2007)

This definition explained how consumers intentionally seek health information and their common seeking situations. This study primarily discusses online health information-seeking behaviors.

Information-seeking behaviors vary predominantly in terms of user demographics. In order to offer tailored services to customers, it is crucial to understand how different groups of people search for health information online. Gary et al. (2005) explored the seeking behaviors of adolescents who came from the United Kingdom and the United States. 157 students who were between eleven to nineteen years old and spoke English were recruited and were classified into twenty-six focus groups. Their attitudes towards the credibility of online health information varied; however, most of them considered online health information as an important health information source. The authors also noted that online communities could facilitate empathy.

In addition to adolescents seeking health information their parents also contributed significantly to finding out the health information for their children. According to a study (Khoo et al., 2008), 52% of the parents had searched health information online for their children. It was no surprise that the most reliable health information sources were doctors and nurses. However, it’s worth noting that 97% of them would trust the local children’s hospital website even though only 20% knew of it before the interview. Based on the results, local hospitals have an obligation to offer online health information to their customers.

In addition to the youth generations, it is also important to study the online health information-seeking behaviors of older people since they are more likely to have chronic diseases and they tend to isolate from society (Tennant et al., 2015). According to a longitudinal study conducted by Flynn, Smith, and Freese (2006) on a group of adults aged 63 to 66 in Wisconsin it was found that one-third of participants had searched health-related information on the Internet. Tennant et al. (2015) sampled 283 baby boomers (born between 1946 and 1964) with an average age of 67.46 years. He noted that three factors were positively correlated to more online health information usage according to the regression analyses: age, education level, and electronic device usage. People of younger age, higher education level, and more digital device usage were more likely to seek health information online.

2.4.2 Topics searched online

According to Fox (2006), 64% of Internet users have searched for specific diseases or medical problems, thus it’s important to understand the health information-seeking behaviors of different groups of patients.

Mayer et al. (2007) examined the information-seeking behaviors of cancer survivors. It was suggested that 67.5% of survivors had searched cancer information online. The Internet and health care providers were the two most used information sources. The survivors who sought cancer information were significantly younger and had higher income and higher education levels than those who didn’t. They were also more likely to have Internet access and other people to look for cancer-related information for them. Park and Park (2014) investigated how Korean Americans sought cancer information since they were less likely to have access to health care and cancer screening tests than common Americans. The text data extracted from an online community, MissyUSA, were analyzed. The results showed that breast cancer was the greatest concern for Korean Americans, followed by cervical and liver cancer.

Diabetes is another frequently studied disease. Shaw and Johnson (2011) sampled fifty-seven patients from the southeastern United States. They found that most patients had sought health-related information online and that 78.5% of them had changed their attitudes toward health after reading the information they found. According to the results, the race factor did not impact online health information-seeking behaviors in terms of the usage of online health information, the usage of social networking sites, and the willingness to discuss health topics online.

2.5. Online health information resources

As the Internet has grown, so too have online information sources grown. 77% of online health information searches began with a search engine such as Google, Bing, or Yahoo (Pew Research Center, 2015). After the initial searches, the consumers were directed to various health-related websites. There are diverse online health information sources to which consumers can obtain access. Cline and Haynes (2001) indicated that most consumers seek health-related information from three online sources: health webpages, including reliable scientific sources and non-reviewed sources; online support groups; and online interaction with health professionals via e-mail. This study discusses two types of online health information resources: social media and online health portals.

2.5.1. Social media

Thanks to the developments of Web 2.0, Internet users interact or collaborate with each other on social media such as Facebook, Twitter, Instagram, LinkedIn, and the list goes on. Merriam-Webster defines social media as “forms of electronic communication (such as websites for social networking and microblogging) through which users create online communities to share information, ideas, personal messages, and other content (such as videos)." Kaplan and Haenlein (2010) defined social media as "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content." With these two definitions, it is clear to see one of the most important characteristics of social media is that users may generate content rather than just browse contents generated by publishers. There are progressively more people tending to search and share health-related information with the rise of social media (Eyrich, Padman, & Sweetser, 2008; McNab, 2009).

2.5.1.1. Online Q&A communities

Online Q&A (question-and-answer) community is a knowledge-sharing platform that allows both professionals and consumers to ask and answer questions. Harper, Raban, Rafaeli, and Konstan (2008) defined the online Q&A community as a website that is "purposefully designed to allow people to ask and respond to questions on a broad range of topics." Shah, Oh, and Oh (2009) suggested that the online Q&A community is a service that allows consumers to present information need using natural language rather than keywords. Beyond ask and answer activities, the authors also noted that Q&A communities provide supplementary services such as making a comment on the answer, rating the answer, and voting the best answers. These supplementary services further improve the interactions between users in a more dynamic atmosphere. The askers not only get answers to their questions but also find out the credibility and quality of the answerers.

There are plenty of Q&A communities out there, such as Yahoo Answers (https://answers.yahoo.com/), Quora (https://www.quora.com), Baidu Knows (https://zhidao.baidu.com), and the list goes on. With its official launch in December 2005, Yahoo Answers quickly drew a large number of users and became one of, if not the most, popular online Q&A communities. Yahoo Answers might be the community that gets the most attention from academia. Gazan (2011) attributed the popularity in academia to two primary factors: the dominant market share and the availability of the research data.

Zhang and Zhao (2013) extracted health-related topics from Yahoo Answers and put them into twelve categories: Cause & Pathophysiology, Sign & Symptom, Diagnosis & Test, Organ & Body Part, Complication & Related Disease, Medication, Treatment, Education & Info Resource, Affect, Social & Culture, Lifestyle, and Nutrient. In order to show the patterns, a multiple-dimensional scaling (MDS) analysis was conducted at both term level and category level. In addition to the MDS analysis, Zhang and Zhao (2014) employed social network analysis to discover the knowledge and patterns from the terms used in the Yahoo Answers. Zhang, Zhao, and Dimitroff (2014) discovered term usage and mined patterns in diabetes-related topics. Oh, Zhang and Park (2016) examined cancer-related questions in Yahoo Answers. A text mining method was adopted, and 420 terms were extracted and classified. It suggested that health professionals should consider askers' demographic, cognitive, social, situational, emotional, and technical contexts for a better understanding of users' information needs. Zhang (2010) conducted a thorough contextual analysis of askers' behaviors, such as the linguistic features of the questions that users formulated, their motivations for asking the questions, the time when the questions were asked, and their cognitive representations of the problem space.

It's not complicated to understand the motivations of askers in the Q&A communities, the reasons of responders are not as evident. Oh (2012) investigated the characteristics and motivations of health-related topics answerers in the Yahoo Answers. The author pre-defined ten motivations. The findings suggested that altruism is the dominant motivation, and enjoyment and efficacy are the other two most important factors that encourage answerers. The understanding of respondents' motivations helps system designers to better incent answerers to interact with the askers.

2.5.1.2. Social network sites

Boyd and Ellison (2007) defined social network sites as web-based services that allow users to:

(1) construct a public or semi-public profile within a bounded system
(2) articulate a list of other users with whom they share a connection
(3) view and traverse their list of connections and those made by others within the system

Examples of social network sites include Facebook, Instagram, Twitter, and so on. According to a survey (Fox, 2011), as of 2010, 62% of adult Internet users used a social network site. Among them, 15% had received health information on the sites, and 9% had joined a health-related group.

Moorhead et al. (2013) conducted a thorough review of what social media brought to health communication using a systematic approach. Six key benefits and twelve limitations were identified at the same time. Although there are many concerns about the quality and reliability of health information on social media, as well as privacy and confidentiality issues, social media is still an important health resource since it offers an efficient information exchange platform for patients, health professionals, and the general public.

Zhang et al. (2013) examined how users communicated with each other in a diabetes group on Facebook. 1352 messages were analyzed. It suggested that although the group members came from all over the world, they managed to overcome the language barriers. Not only was medical information was exchanged among members, but also lifestyle experience. The health group helped members stay positive although living with diabetes.

Twitter also plays an important role in disseminating and obtaining health information. Scanfeld et al. (2010) conducted a content analysis method to categorize antibiotics information shared by users. It concluded that the top three categories were general use of antibiotics, advice on how to use antibiotics, and the side effects. Thackeray et al. (2013) inspected how state health departments shared information on Twitter. They found that state health departments primarily used Twitter as a one-way tool for disseminating information, without interacting with followers. It suggested that state health departments should reconsider their purposes of using Twitter and develop closer relationships with followers.

2.5.1.3. Online social support group

The online social support group not only helps users communicate with each other but also offers health providers and health educators an opportunity to reach target patients (White & Dorman, 2001). An online social support group can be defined as:

> a group of individuals with a common interest or a shared purpose, whose interactions are governed by policies in the form of rules, rituals, or protocols; who have ongoing and persistent interactions; who use electronic communication as the primary form of interaction to support and mediate social interaction and facilitate a sense of togetherness. (Rodgers & Chen, 2005)

Zhang and Yang (2014) explored social support exchange patterns and user behaviors of a smoking cessation social support group by adopting content and network analysis. Two support exchange patterns were identified: initiated support exchange and invited support exchange. The identifications of social support patterns helped the researchers understand how health users help each other which led to designing a more user-friendly and intuitive system. Bender, Jimenez-Marroquin, and Jadad (2011) analyzed the objectives and usages of breast cancer groups on

Facebook. They found that most of the groups were created for fundraising, followed by awareness-raising and product or service promotion. However, there were more users in awareness-raising groups than those in fundraising groups. Chang (2009) scrutinized how Chinese-speaking users communicated with each other in a bulletin board system. A network analysis was conducted to discover the supportive patterns within a psychosis support group. Mo and Coulson (2008) investigated the communication patterns within an online HIV/AIDS support group. It suggested that the informational support and emotional support were the most common support patterns offered in the community.

Some studies focused on investigating how the degree of participation affected users. Pan, et al. (2017) studied an online depression support group. The authors built a network by viewing users as nodes and replies as ties. They found that when they received replies from more diversified sources, or the replies they received were longer in length than the average, the users tended to have more social capital.

Ma and Stahl (2017) investigated the information seeking and sharing behaviors of parents on an anti-vaccine Facebook group. Multimodal critical discourse analysis was undertaken to discover how parents were impacted by vaccine-related scientific research.

2.5.2. Health portal

Luo and Najdawi (2004) noted that a health portal might contain a subject directory that organized the information, a search engine that enabled a keyword search function, a personalized system, and an online user community. The authors further explained how each component helped consumers acquire needed information: the health information catalog facilitated Web editors to organize the health-related information and offered access like hyperlinks to outside health information providers. The search engine functions helped consumers to search for topics within the health portal or on the external websites. The personalization system provided consumers tailored services and a mechanism to edit the user interface. The community function allowed registered users and unregistered users to exchange health-related and unrelated information.

A good example of an online health information portal is MedlinePlus. It is an online health information portal that is operated by the United States National Library of Medicine (NLM). Initially, it provided consumers several functions for better finding needed health information, which included: a list of health-related topics; a dictionary of health-related terms; a database; hyperlinks to health-related organizations; clearinghouses funded by the governments; hyperlinks to news organizations; hyperlinks to libraries; and health provider directories (Miller, Lacroix, & Backus, 2000). These functions greatly helped consumers find information. They indicated that a health portal was a combination of complicated and multifunctioning Web applications.

Keselman et al. (2008) investigated the difficulties encountered when users sought information on MedlinePlus. Twenty participants were recruited to search for information about stable angina. Their key search strategies emerged including verification of the primary hypothesis, narrowing search within the general hypothesis area, and bottom-up search. The findings showed that search skills only helped improve search efficiency. The participants would face difficulties no matter how familiar they were with the Internet search.

Traditional health portals were facing many challenges, from both consumers and Web publishers. Suominen et al. (2009) noted that publishers were facing the issues of duplicated content in the portal. It was difficult to manage the existing content and the content yet to be published. Users of the portal had difficulties finding the content because of the gap between the medical terminology and the terms users usually used. The authors developed a national semantic health portal which consisted of a content aggregation of health ontologies and health-related services, a distributed content creator that let users create semantic content, and a semantic health portal. The newly designed portal helped users find needed content without the bother of duplicate information. Chung et al. (2006) developed an Arabic medical intelligence portal for Arabic-speaking users since non-English-speaking populations were growing rapidly. Many techniques were incorporated into the portal such as meta searching, statistical language processing, and visualization.

2.6 Subject directory and recommendation system

2.6.1. Subject directory

Subject directories are a kind of information organization system that divides information (webpages) into several subject categories based on the characteristics of webpages (Zhang, An & Hong, 2009). This system usually has a hierarchical structure since subject categories can be further divided into sub-categories. Subject directories have been adopted by numerous online portals such as Yahoo.com, health-related websites, and so on.

Chakrabarti et al. (1998) proposed an enhanced categorization algorithm for improving the subject directory. The robust statistical models and a relaxation labeling technique were adopted. The proposed techniques adapted easily to the neighboring documents having known topics. The results show that the proposed classifier outperformed the traditional text classifier.

Yang and Lee (2004) introduced a text mining approach to automatically generate subject directories for websites. The self-organizing map learning algorithm was adopted to create two clusters: a document cluster and a word cluster. These two clusters automatically classified webpages and then generated subjects. The results showed that the proposed methods could automatically produce reasonable subject directories.

Most of the subject directories were developed for all users. However, the need for personalized Web directory services rose. Magoulas, Chen, and Dimakopoulos (2004) found that the cognitive style greatly affected users' information-seeking behaviors, and then developed a flexible interface for subject directories by looking at users' cognitive styles.

2.6.2. Recommendation system

With the developments of the Internet and related technologies, the objectives for satisfying users' information needs became increasingly challenging to achieve. A substantial number of information retrieval systems have incorporated recommendation related methods for fulfilling users' information needs before they performed information search.

Pazzani and Billsus (2007) depicted a common scenario about how the modern recommendation system works: "a system presents a summary list of items to a user, and the user selects among the items to receive more details on an item or to interact with the item in some way." Some examples of recommendation systems were provided, but the definition of the item to be recommended was not articulated in the paper. Ricci, Rokach, and Shapira (2011) put forward the definition of the item as "the general term used to denote what the system recommends to users." In this definition, the item to be recommended could be the commodities sold in E-commerce like Amazon (Linden et al., 2003), the movies from online streaming and media service providers such as Netflix (Zhou et al., 2008) and YouTube (Davidson, 2010), the academic papers from scientific citation indexing databases (Lee et al., 2013), the webpages of Web portals (Sarukkai, 2000), and so on. summarily, from the perspective of the information retrieval field, recommendation systems refer to the systems which satisfy information needs by making suggestions of items to users.

Before talking about the specific recommendation related theories and technologies, some key concepts which laid the foundation of recommendation systems should be clarified beforehand.

Item profiles. In a recommendation system, a profile that represents items' characteristics should be constructed first. The characteristics represented may vary in terms of different items to be recommended. Profiles for the movie, for instance, should include at least four main characteristics: the major actors of the movie, the director, the year the movie was released to the public, and the genre (Leskovec, Rajaraman, & Ullman, 2014). For documents such as articles and text-based webpages, an item profile is represented by a term vector, the weight of the corresponding term can be computed by the term frequency or *tf*idf* value (Pazzani & Billsus, 2007). Some items' characteristics are difficult to extract, such as images. This kind of feature could be obtained from tags that were labeled by the users (Sigurbjörnsson & Van Zwol, 2008).

User-profiles. The user profiles represent the user's preferences in a recommendation system. The preferences are obtained from the interactions of the user and the recommendation system, the interests selected to label oneself, and so on. For instance, in an online movie system, a user profile consists of his rating of movies (Zhou et al., 2008).

Utility matrix. The goal of making a utility matrix is to build the connections between user profiles and item profiles (Leskovec et al, 2014). Table 1 shows an example of a utility matrix of an online movie system. The rows of the matrix indicate the movies while the users are represented by the columns. For example, Vector user1 rated two movies, A and B, on a one to five point scale. One point stands for least favorable, five points means the highest rating. The object of the recommendation system is to decide or predict the blanks in the utility matrix based on item profiles and user profiles. How would user1 like movies C and D? Then the system recommends the movie with a higher estimated rating to the user. As discussed above, the ways to estimate the ratings of the unrated item vary according to the recommendation methods applied.

	USER1	**USER2**	**USER3**	**USER4**
A	5	4	1	
B	4			3
C		3		3
D			5	

Table 1. A utility matrix of an online movie system

2.6.2.1. Content-based methods

Content-based recommendation systems make suggestions based on the properties of items. To be more specific, the recommended items are similar to the items which previously had been ordered or highly rated by users. The similarity is measured by items' characteristics. The major tasks of content-based methods are characteristics extraction, characteristics representation, and similarity measurement between items.

As indicated earlier, the characteristics that could be extracted vary in terms of different types of items. After extracting items' features, the problem becomes how to represent the characteristics to facilitate the similarity measurement between two items. For the documents and text-based webpages, the text is viewed as a “bag of words”, which only regards the frequency of the words while ignoring the grammar and word order. Therefore, the item can be represented as a vector for computing the similarity in the next step. Accordingly, for items such as movies, the characteristics should be conveyed into vectors. In this case, for the characteristics such as actors and genres, the corresponding values could be set as one if the actor is in the movie and zero if not (Leskovec et al., 2014). The final step of content-based methods is to compute the similarity of the items from the database with the items from a user’s profile, and then recommend the items with the highest similarity to the user. The widely used similarity measurement algorithms include Cosine distance and Jaccard distance.

A large number of text-based information systems incorporated content-based methods. Krulwich and Burkey (1996) introduced the thought of content-based methods into a large-scale online information system. The system was designed to make recommendations based on matching the significant phrases extracted from the documents and the interests obtained from users’ self-descriptions. Rather than measuring the similarity between users’ interests and documents’ characteristics, the authors constructed a decision tree, a classifier from machine learning, to categorize the items into two groups, the ones the user might like and the ones the user might not like. Mooney and Roy (2000) also used a machine learning algorithm, a multinomial text model, for book recommendations. The user profile was constructed from users' rating data. Other examples include the news recommendation systems (Lang, 1995), websites recommendation (Pazzani, Billsus, 1997), and so on.

In addition to the text-based information systems, content-based methods can be applied to the other types of information systems. Basu et al. (1998) built a content-based movie recommendation system. This system exploited both rating data and features extracted from movies to make suggestions. The results showed that the new system outperformed the social-filtering method, a technique which only looked into rating data-based systems. Shardanand and Maes (1995) implemented a music information retrieval system, called Ringo, to recommend music albums and artists.

Despite the content-based methods that have been widely accepted and improved, researchers identified three major limitations: Adomavicius &Tuzhilin, 2005). Limited content analysis. In some cases, it was difficult or even impossible to extract features from particular items, such as graphical images, audio streams, and video streams (Adomavicius &Tuzhilin, 2005). Overspecialization (also called serendipity problem). This referred to the problems of the content-based system which could only recommend items that were similar to those which previously interested users. New user problems. Content-based systems were neither efficient nor effective in making suggestions to inexperienced users or to users who were reluctant to rate items.

2.6.2.2. Collaborative filtering methods

On the other hand, collaborative filtering systems rely on the associations between users and items. In other words, a collaborative filtering system recommends items that are bought/read/watched by the “peers” to a user. “Peers” refer to users who share similar interests. Since this method does not require the content information and similarity between items, the drawbacks of limited content analysis and overspecialization from content-based systems can be alleviated or solved. Examples of collaborative filtering methods include traditional collaborative filtering, model based collaborative filtering, and item-based collaborative filtering.

(1). Traditional collaborative filtering

Traditional collaborative filtering methods were also called “memory-based” approaches (Adomavicius & Tuzhilin, 2005; Breese, Heckerman, & Kadie, 1998). These methods make suggestions by looking into a small group of customers who are most like the targeted user. The system first computes the similarities between all the other customers and the targeted user, then categorizes the customers with the highest similarities into a group. Finally, items that are most liked by the group are recommended to the targeted user.

Konstan et al. (1997) implemented an online news recommendation system, GroupLens, which incorporated collaborative filtering methods and provided thoughtful implications. For recommendation systems evaluation it not only assessed the quality of recommendations but also took the cost-benefit analysis into consideration. For example, if a user followed a bad recommendation and went to a cinema, he wasted money and time. But for the news information systems they proposed, a good recommendation helped the users save time searching, but a bad recommendation only gave them a glimpse of it. It suggested that the researchers should be careful to conduct a thorough cost-benefit analysis to eliminate the effects of bad recommendations before designing a system.

Instead of focusing on the personalized recommendation for a single user, Lai (2015) incorporated the knowledge flow mining method to facilitate a group of users who shared similar knowledge and expertise. The author suggested that in a task-oriented environment, the preferences of expert workers who carried some tasks could be extracted for document recommendation. The results indicated that the group-based methods achieved better performance than the personalized recommendation methods in terms of fulfilling a group’s information needs.

The performance of a recommendation system relies on the effectiveness and efficiency of the similarity measurement. The premise of it is that the more accuracy in determining the similar users, the better the recommendations made to the users (Quan et al., 2006). The two most widely accepted measurements were correlation similarity and Cosine similarity (Sarwar et al., 2001). Beyond them, many researchers proposed newly developed or improved algorithms. Sanchez et al. (2008) introduced the mean squared difference (MSD), a statistical method for measuring the average of the squares of the deviations, in user similarity measurement. Compared to Cosine similarity and correlation similarity, the proposed algorithm performed better in providing more accurate predictions and fewer bad predictions. The singularity-based method, a novel algorithm which design was based on the idea that the contribution of an item should be considered differently, was proposed (Bobadilla et al., 2012). The idea of this algorithm, derived from the *tf***idf* algorithm, gave more credit to a term appearing in a document set. Accordingly, the item which was rarely purchased by two users was more important than that commonly bought together. The results suggested the future of the proposed algorithm is promising.

Several limitations of collaborative filtering recommendation systems have been identified, and an abundance of solutions has been established. Since an information system usually contains a huge number of items, the user profile dimension is enormous. As a result, computing the similarity of two users is expensive. Yu et al. (2004) introduced a probabilistic framework for dimensionality reduction. The probabilistic framework identified a subset of user-profiles which keeps the recommendation accuracy while reducing the dimensions dramatically. More recently, Chen et al. (2009) proposed an orthogonal nonnegative matrix tri-factorization method which not only alleviated the computational problem by factorizing the high-dimensional user-item matrix into several smaller matrices but also improved the scalability of the system by clustering users and items at the same time.

(2) Model-based collaborative filtering

Although many researchers tried to improve the performance of traditional collaborative filtering systems by introducing different theoretical frameworks and algorithms, the problems of sparsity and scalability still existed in information systems with an astronomical number of users and items like Amazon. As a result, model-based collaborative filtering methods, an alternative approach to memory-based methods were introduced in both academia and industry.

The key point to a model-based method is to establish a model to categorize the users into several groups in which the users from the same group share similar interests. Therefore, the recommendation system only needs to decide to which group the target user belongs rather than obtaining the similarities between all the users in the system.

Ungar and Foster (1998) introduced three clustering methods (K-means, repeated clustering, and Gibbs) in a movie system for categorizing users based on their preferences. The results showed that the Gibbs sampling method was well suited for modeling. Chee, Han, and Wang (2001) also incorporated a k-means-like clustering method in RecTree.

An alternative modeling method is the Bayesian network, which regards an item as a node in the network. Miyahara and Pazzani (2000) integrated two Bayesian classifiers (the transformed data model and the sparse data model) for making predictions. The results indicated that the transformed data model outperformed the sparse data model in terms of recommendation accuracy. Sarwar (2001) indicated that Bayesian network-based models were more capable since users' preferences were stable.

Apart from clustering and Bayesian Network, many machine learning-related methods have been adopted in classifier modeling. Examples are lazy Bayesian rules (LBR), logistic regression (LR), neural networks (NN), and support vector machine (SVM) (Su et al., 2008). No matter which modeling method is selected, the primary and the most critical requirement is precisely categorizing users or items.

(3) Item-based collaborative filtering

The most critical step for both traditional collaborative filtering methods and model-based methods is to find similar users. Therefore, these two families of methods can be generalized as user-based collaborative filtering methods. However, item-based collaborative filtering methods are proposed to deal with the recommendation problem differently. Instead of focusing on users, this method pays more attention to finding similar items. It does not mean that item-based methods are similar to content-based methods. The difference between them is that item-based methods attempt to match similar items by looking at usage information instead of comparing the characteristics. In other words, the similarity between items was measured based on the rows of the utility matrix (the similarity of users is measured by the columns) which was discussed before, rather than the vectors from item profiles.

A well-known example of item-based collaborative filtering recommendation systems is Amazon. *Touching the Void* is a mountain-climbing book that doesn't have too many attractions since it was published. Many years later, a new book named *Into Thin Air* was published and Amazon noticed that some users bought these books together. Amazon realized that they have some connections and recommended the book *Touching the Void* to those who only bought *Into Thin Air*. As a result, *Touching the Void* became a popular book, even more popular than *Into Thin Air*. This example explains well how item-based methods work and the value of them.

Linden, Smith, and York (2003) further noted that Amazon constructed a product-to-product matrix to present the similarities of all item pairs. Each similarity of an item pair was calculated by Cosine measurement between two vectors, and the dimensions of the vector corresponded to users who had bought or wanted to buy the item. Since the product-to-product matrix could be obtained offline, the item-based systems performed well in a tremendously large data set because of good scalability. Another famous example is YouTube (Davidson et al., 2010). On YouTube, a recommendation happens while a user is watching a video or right after finishing. Therefore, the system recommended similar or related videos of the current video to the user. Similar videos were defined to those that were watched together by the same user.

In academia, the researchers are focusing on improving the performance of the item-based systems. Sarwar et al. (2001) examined three similarity measurements, Cosine measurement, correlation measurement, and adjusted Cosine measurement. Prediction computation methods including weighted sum and regression were tested. The results suggested that adjusted Cosine measurement worked better in item-based systems. The performance of prediction computation algorithms varied as the density of data sets changed. Deshpande and Karypis (2004) investigated the use of conditional probability-based similarity measurement and indicated that item-based methods were much faster and more accurate in recommendation systems than the user-based techniques.

In summary, this chapter reviewed previous studies on the Makov chain analysis method, the transaction log analysis method, online health information needs and seeking behaviors, online information resources, subject directories, and recommendation systems. It’s worth noting that although previous researchers made numerous efforts to improve the users’ information-seeking experience, there were few studies that have incorporate the Markov chain analysis method, the transaction log analysis method, and the inferential statistical method to understand users' traversal patterns. This study proposes a new recommendation approach that incorporates the Markov chain analysis method, the transaction log analysis method, and the inferential statistical method to make recommendations to users at three levels: the topic level, the article level, and the Q&A item level on a health portal.

Chapter 3. Research Methodology

3.1 Introduction

This study investigates health users' traversal patterns on online health portals, to see whether data extracted from the transaction log can recommend related information to users.

The research objectives of this study are focused on the user's traversal activities on a health portal. The research population is all the users and their traversal data in a health portal with a subject directory. Since there are numerous health portals out there, a sampling strategy is proposed to select a representative health portal. The transaction log of that health portal should be accessible and available to researchers for extracting users' traversal data.

A transaction log records users' activities on the portal, which includes IP address, timestamp, visiting point, browser, and referrer. This data can help draw users' traversal paths on the portal. The traversal paths depict how users traverse on the portal, from an article to another article, from a Q&A item to an article, from a topic to a Q&A item, from a topic to another topic, and from a topic to an article. Those traverses indicate that there are relationships among Q&A items, articles, and topics that are connected by the traversal paths.

The Markov chain method is employed to model those traversal paths. To construct the transition matrix for the Markov chain analysis, textual data collected from the articles are also used to calculate the similarities between articles and topics.

3.2 Data collection

The data collected in the study includes two parts. The first part is the users' traversal data extracted from the transaction log. The second part is the textual data gathered from the articles in the HealthLink portal.

3.2.1 Sampling strategy

The purpose of the study is to investigate the health users' traversal data on online health portals for recommending related items, thus the sampling strategy is to find an appropriate health portal. Selecting a suitable sample for the study is essential to achieving the research's objectives since the sampling process largely determines the quality of the research in terms of reliability, validity, and generalizability. In this study, the sampling strategy should meet the requirements in design, traffic, representation, and availability.

Given the requirements, the Medical College of Wisconsin's HealthLink portal is selected to serve as the sample in the study. HealthLink provides consumer health information services to its patients and the community. It incorporates a health subject directory that provides health consumers with an information browsing mechanism in addition to the query search. The health users can find information using their preferred search method.

HealthLink's subject directory is a three-level hierarchical structure system. The root is called "browse by topic", where users start browsing articles. The nodes on the directory branches are topics. In the directory, there are forty-seven topics/nodes. These forty-seven subjects are presented in Table 2 in alphabetical order. Each node contains a variety of related articles and Q&A items. Users can traverse within the nodes and articles at will. Figure 7 shows the hierarchical structure of the subject directory. This figure indicates that the portal is well-designed for facilitating users' information seeking by classifying the articles into a pre-defined category in which users can easily find topics in which they are interested.

Initial	Category
A	Aging Allergies Alternative medicine Arthritis
B	Back problems Brain nervous system
C	Cancer Children's health Cholesterol Clinical trials
D	Diabetes Digestive disease Drugs medications
E	Emergency medicine Endocrine system Environmental health Eye care
F	Feet
G	Genetics
H	Hearing disorders Heart disease High blood pressure
I	Immune disorders Immunization Infections
K	Kidney disease
L	Liver
M	Men's health Mental health Musculoskeletal
N	Neurological disorders Nutrition and herbs
O	Occupational health Organ transplants
P	Pain; Physical medicine Preventive medicine Public health
R	Respiratory
S	Safety Skin diseases Sports medicine

T	Travel medicine
V	Vitamins
W	Weight control Wellness lifestyle Women's health

Table 2. The forty-seven nodes in the Medical College of Wisconsin's HealthLink portal

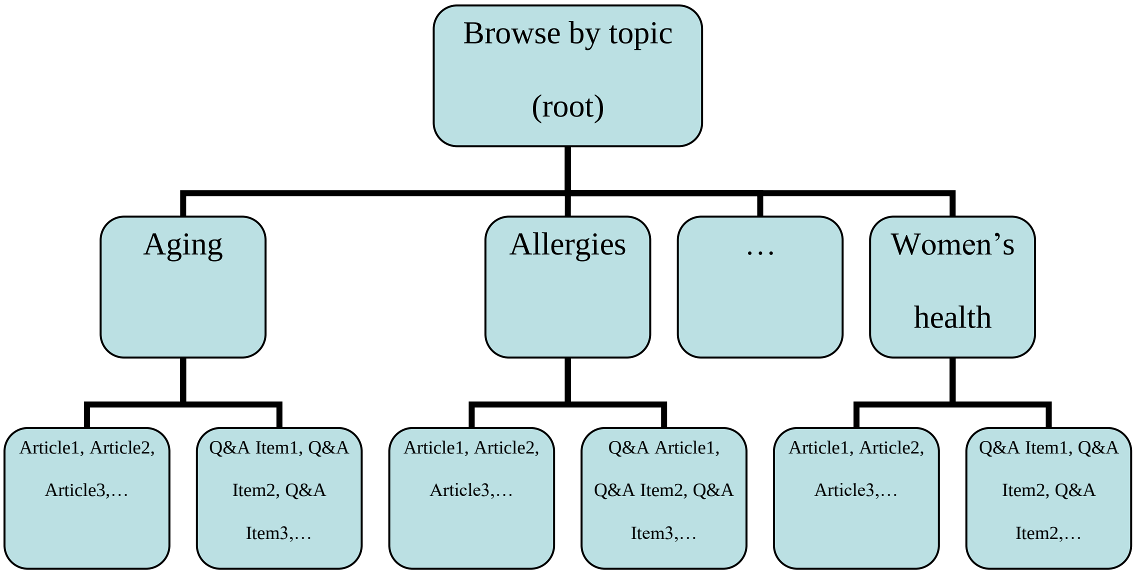


Figure 7. The hierarchical structure of the subject directory

There are 2368 articles from forty-seven subjects that provide sufficient health-related information to the users; 636 of them are Q&A items. It was visited more than 266,000 times and 1,675,000 pages have been viewed per month. Based on the analysis of Internet Protocol (IP) address, the users come from around the world but most of them are from North America. It could be concluded that the users of the portal are representative enough, and the traffic of the system is sufficient to be studied.

3.2.2 Transaction log data collection procedure

In the HealthLink portal, all users' activities are recorded in the transaction log. In the study, the activities from bots and spams were filtered by using a transaction log analyzing tool, WebLog Expert. Only activities conducted by users were kept for analysis. A typical piece of transaction log data can be read as follows.

```
65.214.44.193 - - [01/Jan/2006:00:00:40 -0600] "GET /article/956024919.html
HTTP/1.0" 200 17693 "-" "Mozilla/2.0 (compatible; Ask Jeeves/Teoma;
+http://sp.ask.com/docs/about/tech_crawling.html)"
```

This data can be further divided into six parts.

IP address: 65.214.44.193. The first part of data, the IP address indicates from where the user came. It can be used to identify users.

Timestamp: `01/Jan/2006:00:00:40`. It shows when the user conducted this activity.

Method: Get. The HTML method used by the user. There are two types of methods, the get method and the post method. In this case, the user used the get method.

Visiting point:`/article/956024919.html`. It indicates where the user on the portal at the current state is. It's crucial for later analysis.

Browser: *`Mozilla`*. This information shows which browser the user used.

Referrer: *`http://sp.ask.com/docs/about/tech_crawling.html`*. It shows from where the user came.

The major reason for collecting transaction log data is to identify the traversal paths. A traversal path is presumably associated with a unique user. However, it's nearly impossible to identify a user identity by using the transaction log data. Therefore, sessions are used to identify traversal paths in the study. Data extracted from IP address and timestamp are used to identify user sessions; visiting pointing and referrer are used to find paths. Data entries that share the same IP address and all the traversal activities occurring in a two hour timeframe are considered as a user session. It's worth noting that a session is not equal to a user since an IP address might be used by different users. There are two reasons for using two hours as the threshold for one session. If the timeframe is too short, it may not be able to identify enough traversal paths. If the timeframe is too long, a session may not represent the same user.

After session segmentation, each user will receive a user ID and each subject node on the health information portal will receive a subject ID. Visiting times are kept.

Each visit of a user on a node of the navigation system is organized as follows:

$$V(i) = \{UID, CSID, NSID, VT\} \quad (2)$$

In Equation 2, *i* is the visit number. *V(i)* is a visit activity. *UID* stands for user *ID, CSID* for current subject node *ID or start subject node ID*, *NSID* for next subject node *ID or end subject node ID*, and *VT* for visiting time.

If the corresponding *CSID* is empty, it suggests that there is no jump between two subject nodes. It is either the end of a user visit path or an isolated visit of a user. Therefore, it is discarded because it does not contribute to visit activity analysis.

As a result, all visit activity data on the navigation system are shown:

$$VAD = \{V(i)\}, i \leq N \quad (3)$$

In Equation 3, *N* is the number of all user visit activities on the navigation system.

The *VAD* can form a list which has 4 columns *UID*, *CSID*, *NSID*, and *VT*. Items on the *VAD* list are ranked against *UID* first. Consequently, all visit activities of a user are grouped.

Each user activities group is also shown as:

$$UV(j) = \{CSID, NSID, VT\}, 1 \leq j \leq M \quad (4)$$

Here *UV(j)* is all visit activities of user *j* and *M* is the number of users in the transaction log.

We also have:

$$VAD = \{UV(j)\}, 1 \leq j \leq M \quad (5)$$

The second-round ranking is conducted within each user group (*UV*(*j*)). In other words, the ranked list tells a sequence of his/her visits on the navigation system based on his/her visiting time.

In other words, a user's visit individual activity is defined as:

$$UVIA(j, k) = \{CSID, NSID, VT\}, 1 \leq j \leq M, 1 \leq k \leq L \quad (6)$$

In Equation 6, *UVIA* (*j*, *k*) is the individual activity *k* for user *j*, and *L* is the number of all activities of that user. An activity of a user, in this case, refers to a move or jump from one subject node (*CSID*) to another subject node (*NSID*) on the navigation system.

It is clear that all activities of a user are defined as:

$$UV(j) = \{UVIA(j, k)\}, 1 \leq j \leq M, 1 \leq k \leq L \text{ (7)}$$

It is necessary to define a user visit session because a user can do multiple tasks and make multiple visits on the same portal. In addition, multiple users can use the same public computer as a computer in a public library or a public lab to access the portal. It is apparent that a user visit session consists of a group of consecutive visit activities.

Assume user *j* has a group of visit activities described in Equation (7).

If *VT(UVIA(j, k))* is defined as the time of visit activity *k* for user *j*, then the following equation is always true.

$$VT\big(UVIA(j,k)\big) < VT\big(UVIA(j,k+1)\big), 1 \le j \le M, 1 \le k \le L-1 \text{ (8)}$$

The following visit activities in Equation 9 are defined as a visit session if Equation 10 is met.

$$VS(j,s) = \{UVIA(j,k_1), UVIA(j,k_2), \dots, UVIA(j,k_s)\}, j \le M, 1 \le k_s \le L \text{ (9)}$$

$$VT\left(UVIA(j,k_{y+1})\right) - VT\left(UVIA(j,k_y)\right) \le C, 1 \le j \le M, 1 \le k_y \le L-1 \text{ (10)}$$

In Equation 10, *C* is a predefined threshold. According to a study, *C* was equal to twenty minutes (Wolfram, Wang, & Zhang, 2009). Equation 10 suggests that if the time difference between two adjacent activities is larger than the constant, it is regarded as the border of two visit secessions. In other words, if the duration that a user stays in one subject node is longer than a constant, the subject node is the end of a session.

$$UV(j) = \{VS(j,s)\}, 1 \le j \le M, 1 \le s \le P \text{ (11)}$$

In Equation 11, *s* is the number of all visit sessions for user *j*.

In summary, on the resultant *VAD* list, all users are identified, all visiting activities of a user are specified in the first-round ranking against user *ID*s, and the visit sessions of a user on the health information portal are determined after the second round ranking against visiting time.

3.2.3 Textual data collection procedure

In this study, textual data from articles were extracted to conduct vector space model analysis. All 2,368 articles in the health portal were collected, only title and text body information were extracted for the text analysis. All other information, such as ads, hyperlinks, and other HTML information were removed because they did not contain meaningful information for content analysis. Each article was segmented into words. Then a stop-word list was used to filter words that are not content bearing such as “the”, “and”, “or”, etc. Finally, each article was converted to a vector using the term frequency-inverse document frequency (Tf-idf) method, which amplifies the influence of uncommon words but reduces the influence of commonly used words in the document set.

3.2.4 Relevance judgment data collection

To answer the proposed research questions, experts were recruited to judge the relevance of the recommended items with the items of interest. Five experts with M.D. (Doctor of Medicine) were recruited to make the judgment. The recommended items were judged based on a ten-point rating scale: one for the least relevant and ten for the most relevant. After collecting the relevance scores, the average relevance scores were calculated. In order to keep the consistency of the relevance data, the Krippendorff's alpha test was introduced to ensure the inter-coder reliability. The Krippendorff's alpha coefficient was a statistical measurement to test the inter-coder agreement (Hayes & Krippendorff, 2007). It has been adopted by various studies on relevance judgment. It’s not affected by the sample sizes or missing values, which means it can be used when there are more than two raters.

3.3 Validity and reliability

3.3.1. Validity

A study has validity, also called internal validity, “if it produces a single, unambiguous explanation for the relationship between two variables” (Gravetter & Forzano, 2015). General threats to the validity of a study include environment variables, assignment bias, history, maturation, instrumentation, testing effects, and statistical regression (Gravetter & Forzano, 2015).

In this study, validity is ensured by the following methods. Firstly, the completeness of traversal data is ensured. All user activities of the HealthLink portal are recorded in the transaction log. With all data extracted from the transaction log, users’ traversal paths can be easily drawn. Secondly, five experts are recruited to evaluate the proposed recommendation system. To better ensure the validity, a consistency test among evaluation results of the five experts is performed to make sure their evaluation results are reliable and consistent. Thirdly, while conducting textual data analysis, not only titles but also full texts of articles and Q&A items are used to make sure the results of the analysis are more precise than just using titles. Lastly, the researcher uses twelve transaction sublogs in the study. Each sublog records the users’ activities each month throughout a full year. In other words, the transaction logs cover various users’ activities in a full year, which is sufficient to represent the population.

3.3.2. Reliability

Bryman (2015) suggests that a study is reliable if the results of the study are repeatable under similar circumstances. The reliability of this study is ensured in two ways. First, users’ traversal patterns may vary largely at different periods of the year. The researcher uses a transaction log that records users' activities for a whole year, which ensures the consistency and sustainability of the study.

Second, during the recommendation system evaluation process, five experts are recruited to test whether the proposed system can efficiently recommend items to users. The intraclass correlation coefficient (ICC) reliability formula is adopted to test the inter-coder reliability of five experts. In this study, a fair agreement (0.4) among five experts is accepted to ensure the reliability of the study (Cicchetti, 1994).

An intraclass correlation coefficient is a statistical method that can be used in many fields, which include medical, biological, psychological, and many others. It describes the correlations within a group of data, rather than the correlations between two groups of data with different structures. Therefore, when different observers measure the same objects, the ICC can be used to assess the consistency of quantitative measurements made by them (Koo & Li, 2016). In the study, the ICC is used to assess the degree of agreement among raters. When multiple raters measured the same objects, the conformity or consistency among the raters can be ensured if a substantial agreement of ICC is achieved. The five experts were recruited to rate the relevance scores of the returned results to a given item. Therefore, the intraclass correlation coefficient reliability formula is suitable to ensure the consistency of the relevance scores rated by the different experts.

3.3.3 Generalizability

Generalizability, also called external validity, stands for “the extent to which we can generalize the results of a research study to people, settings, times, measures, and characteristics other than those used in that study” (Gravetter & Forzano, 2015).

In this study, the generalizability is ensured by selecting a representative sample. As discussed above, the Medical College of Wisconsin’s HealthLink portal is selected to serve as the sample in the study. It’s a comprehensive health portal that provides varied health information to users. A subject directory is incorporated into the portal that enables users to browse related articles and Q&A articles. And last, the health portal records users’ activities in transaction logs. Therefore, future studies can incorporate the proposed system in other information systems in different fields that have a hierarchy subject directory to organize information, have full-text articles to convey information, and have a transaction log to record users’ activities.

3.4 Markov chain analysis

3.4.1 Markov chain theory

The Markov chain method, named after the Russian mathematician Andrey (Andrei) Andreyevich Markov (1856-1922), is a special stochastic process which is “memoryless” (Marcellin & Fischer, 1990). In the study, the first step of the Markov chain analysis is to create transition matrices.

3.4.2 Creation of the related matrices

After all traversal data in the system are collected, several transition matrices are generated. These matrices indicate the traversal activities between articles and the traversal activities between topics. The generation of matrices is crucial and vital for Markov chain analysis.

Creation of original frequency matrices

(1) Article-Article Matrix (*AAM*): traverses between all of the articles.

$$AAM = \begin{pmatrix} A_{1,1} & \cdots & A_{1,r} \\ \vdots & A_{i,j} & \vdots \\ A_{r,1} & \cdots & A_{r,r} \end{pmatrix} (13)$$

Here *r* is the number of articles in the system including the general articles published by the doctors. $A_{i,j}$ is a cell in the *AAM* matrix, which indicates the number of traverses from article *i* to article *j*.

(2) Q&A item-Q&A item Matrix (*QQM*): traverses between the Q&A articles.

$$QQM = \begin{pmatrix} Q_{1,1} & \cdots & Q_{1,s} \\ \vdots & Q_{i,j} & \vdots \\ Q_{s,1} & \cdots & Q_{s,s} \end{pmatrix} (14)$$

Here, s is the number of Q&A articles in the system. $Q_{i,j}$ is a cell in the *QQM* matrix, which indicates the number of traverses from Q&A article *i* to article *j*.

(3) Topic-Topic Matrix (*TTM*): traverses between topics.

$$TTM = \begin{pmatrix} T_{1,1} & \cdots & T_{1,47} \\ \vdots & T_{i,j} & \vdots \\ T_{47,1} & \cdots & T_{47,47} \end{pmatrix} (15)$$

Here, forty-seven is the number of topics in the system. $T_{i,j}$ is a cell in the *TTM* matrix, which indicates the number of traverses from topic *i* to topic *j*. The number of traverses between the two nodes is summarized by the total traverse counts between the articles from the topic *i* and the article from topic *j*.

(4) Article-Topic Matrix (*ATM*).

$$ATM = \begin{pmatrix} A_{1,1} & \cdots & A_{1,2368} \\ \vdots & A_{i,j} & \vdots \\ A_{47,1} & \cdots & A_{47,2368} \end{pmatrix} (16)$$

Here, 2368 is the number of articles in the system, and forty-seven is the number of topics in the system. $A_{i,j}$ is the cell in the *ATM* matrix, which refers to whether the article *j* belongs to the topic *i*. If $A_{i,j}=0$, it indicates that the article *j* doesn't belong to the topic *i*, Whereas $A_{i,j}=1$ suggests that topic *i* contains article *j*.

Conversion of original frequency matrices to transition matrices

In order to apply the Markov chain analysis, the original frequency matrices are converted to the transition matrices, which are the probability matrices. The sum of each row is equal to 1. Using Equation (17) as an example, the sum of the first row equals 1: $\sum_1^j A_{1,n} = 1$.

(5) Article-Article Transition Matrix (*AATM*): traverses between all of the articles.

$$AATM = \begin{pmatrix} A'_{1,1} & \cdots & A'_{1,r} \\ \vdots & A'_{i,j} & \vdots \\ A'_{r,1} & \cdots & A'_{r,r} \end{pmatrix} (17)$$

Here, *r* is the number of all articles to be analyzed including the general articles published by the doctors. $A'_{i,j}$ is a cell in the *AATM* matrix, which indicates the probability of transition from article *i* to article *j*.

(6) Q&A item-Q&A item Transition Matrix (*QQTM*): traverses between the Q&A articles.

$$QQTM = \begin{pmatrix} Q'_{1,1} & \cdots & Q'_{1,s} \\ \vdots & Q'_{i,j} & \vdots \\ Q'_{s,1} & \cdots & Q'_{s,s} \end{pmatrix} (18)$$

Here, s is the number of Q&A articles to be analyzed. $Q'_{i,j}$ is a cell in the *QQTM* matrix, which indicates the probability of transition from Q&A article *i* to the article *j*.

(7) Topic-Topic Transition Matrix (*TTTM*)

$$TTTM = \begin{pmatrix} T'_{1,1} & \cdots & T'_{1,47} \\ \vdots & T'_{i,j} & \vdots \\ T'_{47,1} & \cdots & T'_{47,47} \end{pmatrix} (19)$$

Here, forty-seven is the number of topics in the system. $T'_{i,j}$ is a cell in the *TTTM* matrix, which indicates the probability of transition from the topic *i* to the topic *j*. The number of traverses between two nodes is summarized by the total traverse counts between the articles from the topic *i* and the topic *j*.

Modification of the transition matrices

The transition matrices need to be modified before conducting the Markov analysis. The major reason for doing this is that the *AATM* and *QQTM* matrices are sparse matrices due to the lack of traversal data between articles. On average, 60% of the entries are empty. For instance, if there are only three non-zero values in one column in the *AATM* matrix, it means that only three articles have traversal activities between that corresponding article. That being said, it suggests that users will only visit these three articles and there is no chance to visit others in the future. However, it's wrong to suggest that in the real world. Therefore, it's crucial to fill up empty entries with positive values. In this study, the techniques adopted to fill up the empty entries are brought up by Page, Brin, Motwani, and Winograd (1999). They supposed that there was 85% of chance that a random user would follow the existing route, and there was 15% of chance that the user would take a non-existing route.

Let *T* be the transition matrix that is calculated by traversal data, let *S* be another transition matrix that is calculated by using Cosine similarity, the modified transition matrix G can be obtained by combining matrices S and T through Equation (20):

$$G = \alpha T + (1 - \alpha)S \qquad (20)$$

In Equation (20), α=15%. It suggests that a random user will follow the existing traverse path obtained by using traversal data in 85% of chance, and there is 15% chance that this user will visit a random article or topic on the portal. By conducting this modification, it not only converts the sparse matrix into a positive matrix, but also makes the users' traversal activities on the portal more reasonable. Therefore, the final transition matrices can be obtained:

(8) Modified Topic-Topic Transition Matrix (*MTTTM*)

$$MTTTM = \begin{pmatrix} T''_{1,1} & \cdots & T''_{1,47} \\ \vdots & T''_{i,j} & \vdots \\ T''_{47,1} & \cdots & T''_{47,47} \end{pmatrix} = \alpha\mathrm{TTTM} + (1-\alpha)STTM=$$

$$\begin{pmatrix} \alpha T'_{1,1} + (1-\alpha)S_{1,1} & \cdots & \alpha T'_{1,47} + (1-\alpha)S_{1,47} \\ \vdots & \alpha T'_{i,j} + (1-\alpha)S_{i,j} & \vdots \\ \alpha T'_{47,1} + (1-\alpha)S_{47,1} & \cdots & \alpha T'_{47,47} + (1-\alpha)S_{47,47} \end{pmatrix} \quad (21)$$

In Equation (21), STTM stands for Similarity Topic-Topic matrix. Each entry represents the Cosine similarity between two topics. $S_{i,j}$ is the value of Cosine similarity between topic *i* and topic *j*. Each entry of the *MTTTM* is the transition probability between two topics. $T''_{i,j}$ is the transition probability between topic *i* and topic *j*.

(9) Modified Article-Article Transition Matrix (*MAATM*)

$$MAATM = \begin{pmatrix} A''_{1,1} & \cdots & A''_{1,r} \\ \vdots & A''_{i,j} & \vdots \\ A''_{r,1} & \cdots & A''_{r,r} \end{pmatrix} = \alpha\mathrm{AATM} + (1-\alpha)SAAM=$$

$$\begin{pmatrix} \alpha A'_{1,1} + (1-\alpha)S_{1,1} & \cdots & \alpha A'_{1,r} + (1-\alpha)S_{1,r} \\ \vdots & \alpha A'_{i,j} + (1-\alpha)S_{i,j} & \vdots \\ \alpha A'_{r,1} + (1-\alpha)S_{r,1} & \cdots & \alpha A'_{r,r} + (1-\alpha)S_{r,r} \end{pmatrix} (22)$$

In Equation 22, *SAAM* stands for Similarity Article-Article matrix, each entry represents the Cosine similarity between two articles. $S_{i,j}$ is the value of Cosine similarity between article *i* and article *j*. In *MAATM*, $A''_{i,j}$ is the transition probability between article *i* and article *j*.

(10) Modified Q&A item-Q&A item Transition Matrix (*MQQTM*)

$$MQQTM = \begin{pmatrix} Q''_{1,1} & \cdots & Q''_{1,s} \\ \vdots & Q''_{i,j} & \vdots \\ Q''_{s,1} & \cdots & Q''_{s,s} \end{pmatrix} = \alpha\mathrm{QQTM} + \ (1-\alpha)SQQM =$$

$$\begin{pmatrix} \alpha Q'_{1,1} + (1-\alpha)S_{1,1} & \cdots & \alpha A'_{1,s} + (1-\alpha)S_{1,s} \\ \vdots & \alpha A'_{i,j} + (1-\alpha)S_{i,j} & \vdots \\ \alpha Q'_{s,1} + (1-\alpha)S_{s,1} & \cdots & \alpha A'_{s,s} + (1-\alpha)S_{s,s} \end{pmatrix} (23)$$

In Equation 23, *SQQM* stands for Similarity Q&A item Q&A item matrix, each entry represents the Cosine similarity between two Q&A items. $S_{i,j}$ is the value of Cosine similarity between Q&A item *i* and Q&A item *j*. In *MQQTM*, $A''_{i,j}$ is the transition probability between Q&A item *i* and Q&A item *j*.

In a summary of the matrices, matrix *MTTTM* is used to recommend the topics, which aims to answer RQ1 and the related sub-questions. Matrix *MAATM* is used to recommend articles, which aims to answer RQ2 and the related sub-questions. Matrix *MQQTM* is used to recommend Q&A items, which aims to answer RQ3 and the related sub-questions.

3.4.3 Analysis of the topic matrix

As discussed above, one of the most critical attributes of the Markov chain is that, if the transition matrix is regular, then there is a vector *V* which satisfies Equation (12) with an arbitrary vector *v*. The vector *V* can also be called a stationary distribution. In this study, the first step is to find stationary distributions for Markov chains.

In equation (20), the 47×47 square matrix $MTTTM$ represents the transition probabilities from topics to topics. The Markov chain analysis is conducted to examine if there is a stationary distribution V_{topic} that can fit the Equation (21):

$$\lim_{n\to\infty} v_{topic} \times MTTTM^n = V_{topic} \quad (24)$$

In Equation 24, v_{topic} is an arbitrary 47^{th} order vector, which serves as the initial state of the Markov chain. V_{topic} is a stationary 47^{th} order vector, which represents the final or stationary state of the Markov chain. V_{topic} is represented in Equation (25):

$$V_{topic} = \{p_1, p_2, p_3, p_4, \dots \dots, p_{46}, p_{47}\} \quad (25)$$

In this Equation, p_n represents the final/stationary probability of staying at Topic *n*, $\sum_1^{47} p_n = 1$. Therefore, V_{topic} is the vector that indicates the possibility of a user staying at each topic at the final state. This vector can be used to rank the importance of all the topics based on the probability distribution.

The ranking of the topics can help us re-organized the subject directory structure of the health portal. The highly ranked topics can be placed at the very beginning of the more obvious position on the homepage to replace the topics at the end of the list.

3.4.4 Analysis of the article matrix

The analysis of the article matrix focuses on the *MAATM*. However, there are a total of 2368 articles on the portal. It is time-consuming to conduct a study to such an extent. In this study, 15 articles from 3 topics (*Allergies/Asthma*, *Kidney*, and *Diabetes*) are selected to conduct the analysis. Each topic selects 5 articles. Therefore, 15 *MAATM* matrices are established based on the traversal data. The portal contained 2,368 articles in total. In theory, the article transition matrix can be generated based on all the articles. However, this study generated thirty small transition matrices for the Markov chain analyses instead. There are two main reasons behind that. The first one is the computational cost for the huge matrix. If a higher Markov chain were used, the computational cost would increase dramatically. The second is the dynamic nature of the Markov chain method. The second reason is also related to the first one. The recommendation lists can be generated dynamically based on the traversal paths collected from recent transaction log. If a Markov chain analysis must be done whenever the huge transition matrix was updated, it would bring a huge computational cost. Therefore, with small transition matrices, the computational cost was accepted. It is also true when a new article is added to the portal. In the future research, the effectiveness comparison between large matrix and multiple smaller matrices will be conducted.

In this step, an experiment is conducted to find if there is a vector $V_{article}$ that satisfies Equation (24):

$$\lim_{n\to\infty} v_{article} \times MAATM^n = V_{article} \ (26)$$

In Equation 26, $v_{article}$ is a random r^{th} order vector, which serves as the initial state of the Markov chain. $V_{article}$ is a stationary r^{th} order vector, which represents the final or stationary state of the Markov chain. $V_{article}$ is represented in Equation (25):

$$V_{article} = \{p_1, p_2, p_3, p_4, \dots \dots, p_{r-1}, p_r\} \quad (25)$$

In Equation (25), p_n represents the final/stationary probability of staying at Article *n*, $\sum_1^r p_n = 1$. Therefore, $V_{article}$ is the vector that indicates the possibility of a user staying at each article at the final state. This vector can be used to indicate the importance of articles.

3.4.5 Analysis of the Q&A article matrix

The *Q&A articles* consist of an important part of the health portal. Another Markov chain analysis is necessary to find if there is a stationary vector $V'_{article}$ that satisfies the following Equation:

$$\lim_{n \to \infty} v'_{article} \times MQQTM^n = V'_{article} \quad (27)$$

In Equation (27), $MQQTM$ is a $s \times s$ square matrix that was constructed to represent the traversal patterns of the s *Q&A* articles in any topic. $v'_{article}$ is an arbitrary s^{th} order vector, which serves as the initial state of the Markov chain. $V'_{article}$ is a stationary s^{th} order vector, which represents the final or stationary state of the Markov chain. $V'_{article}$ is represented in Equation (28):

$$V'_{article} = \{p_1, p_2, p_3, p_4, \dots \dots, p_{s-1}, p_s\} \; (28)$$

In Equation (28), p_n represents the final/stationary probability of staying at Article *n*, $\sum_1^s p_n = 1$. Therefore, $V'_{article}$ is the vector that indicates the possibility of a user staying at each article at the final state. This vector can be used to indicate the importance of all the Q&A articles under any category.

3.5 Generation of the recommendation lists

After analyzing the modified matrices, the next step is generating recommendation lists for users. The recommendation lists are presented in the following three scenarios.

3.5.1 Topic recommendation lists

The topic recommendation lists are offered when users are browsing the topic page. Figure 8 shows the *Allergies/Asthma* topic page view. This page lists articles that belong to that topic. A topic recommendation list is offered on the right of the article list.

The list can be generated through the following two steps:

Find the most relevant topics. Use the Topic *Allergies/Asthma* as an example. Forty-six Cosine similarity calculations are conducted for the rest of the forty-six topics. These forty-six similarity values are ranked from the largest value to the smallest. Fifteen most similar topics are selected from forty-six topics. Each topic vector was calculated based on the average vector of all the articles from the topics. Using Topic “Cancer” for an example, it included 306 articles, 306 vectors representing these 306 articles were generated by using the cosine method. The vector of the Topic Cancer was then generated by obtaining the average vector from these 306 article vectors.

Re-rank the topics. Use the value of p_i in vector V_{topic} to re-rank the fifteen most similar topics and select 10 most relevant topics in terms of the value of p_i. Then present the topic recommendation list which includes these 10 topics to the users.

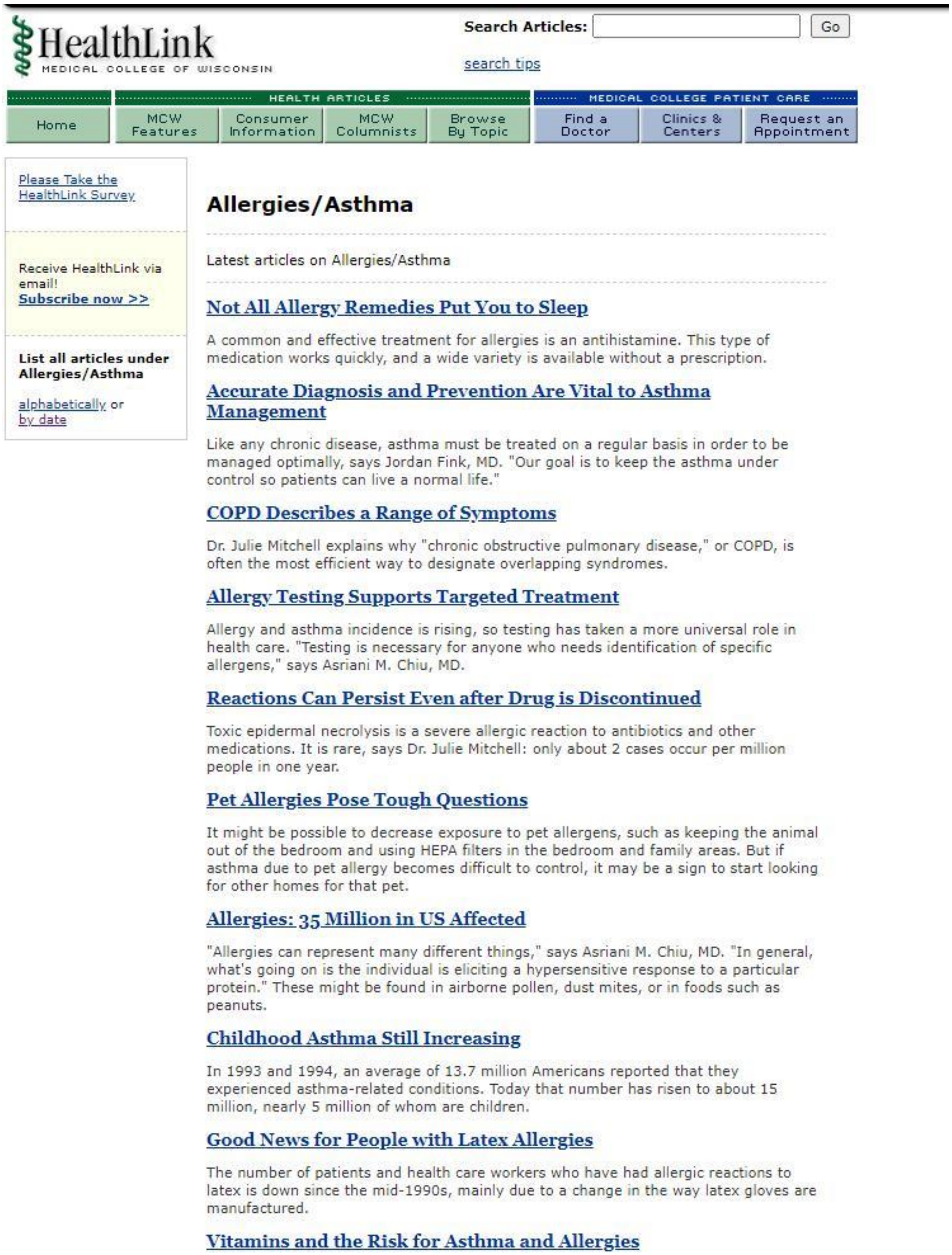


Figure 8. The topic page view.

3.5.2 Article recommendation lists

The article recommendation lists are offered when the users are browsing the article page. Figure 5 above shows an article of the *Allergies/Asthma* topic page view. This page only lists related topics but no recommended articles. An article recommendation list is offered on the right of the content.

The list can be generated through the following two steps:

Find the pool of the recommended articles to the anchor article. Use the article from Figure 5 as an example. Extract keywords from the title and the keywords feature. In this case, the keywords are nasal, sinus, and polyps. Find the articles that contain these keywords by using the Boolean search. The pool contains 20-30 related articles.

Re-rank the articles in the pool. Use the value of p_i in vector $V_{article}$ to re-rank these articles and select 15 articles with the largest values p_i. Then the article recommendation list which includes these 15 articles is offered to the users.

3.5.3 Q&A item recommendation list

Like the article list, the Q&A item recommendation lists are offered when the users are browsing the article page. Figure 6 shows a Q&A item about the *Alternative Medicine* topic page view. This page only lists related topics but no recommended Q&A items. A Q&A item recommendation list is offered on the right of the content.

Similarly, the list can be generated through the following two steps:

Find the pool of the recommended Q&A items. Use the Q&A item from Figure 6 as an example. Extract keywords from the title and the keywords feature. In this case, the keywords are allergy, remedy, and sleep. Find the Q&A items that contain these keywords by using the Boolean search. The pool contains 20-30 related Q&A items.

Re-rank the Q&A items. Use the value of p_i in vector $V'_{article}$ to re-rank these items and select 15 Q&A items with the largest values p_i. Then the Q&A item recommendation list which includes these 15 items is offered to the users.

In summary, thirty-three recommendation lists are generated in the study, three for the topic recommendation, fifteen for the article recommendation, and fifteen for the Q&A item recommendation.

3.6 Evaluation analysis and inferential analysis

3.6.1 Evaluation analysis

Thirty-three recommendation lists were evaluated and ranked by the experts to test whether the proposed recommendation system can effectively recommend related information to users at the three levels. Five experts in the health domain were recruited to conduct the evaluation. Figure 9 shows an evaluation form for an article recommendation list. On the left-hand side of the form, it's the title and full text of an article. On the right-hand side, it's the recommendation list for this article which includes twenty articles. The items on the recommendation lists were presented to experts in a random order. Titles and full texts of the twenty recommended items were offered separately. The experts were asked to judge the recommended items in terms of the relevance to their anchor article and then gave a relevance score for each of the items. The recommended items were judged based on a ten-point rating scale: One for the least relevant and ten for most relevant. After collecting the relevance scores, the average relevance scores were calculated and the items were ranked according to their scores. The topic evaluation forms, and the Q&A item evaluation forms were designed the same. Among the thirty-three evaluation forms there are three evaluation forms for the topic recommendations, fifteen evaluation forms for the article recommendations, and fifteen evaluation forms for the Q&A item recommendations.

Nutrition During Childhood Cancer

Good nutrition is an important part of your child's treatment. In general, your child's normal diet should be continued during cancer treatment unless your physician gives you a special one. A few diet hints are listed below:

- Build meals around your child's favorite foods. Variety is not as important as intake.
- Small, frequent meals and snacks are attractive to most children. You can freeze portions of a favorite dish and serve them when desired.
- Smaller bites and frequent sips of water, milk, or other unsweetened drinks will make chewing and swallowing easier.
- Avoid empty calorie foods such as soft drinks, chips, and candy that can reduce your child's appetite without providing nutrients. By contrast, milkshakes, yogurt, fruit, juices, or instant breakfasts provide extra calories and protein.
- Some types of chemotherapy may temporarily alter your child's sense of taste. Well-seasoned foods such as spaghetti, tacos, and pizza may seem especially good at times. Sometimes adding extra salt or sugar, or using less, may make foods taste better. However, because of fluid retention, patients on cortisone drugs should limit salt in their diets.
- A decrease in appetite is common to some types of chemotherapy. But this must be countered with an increase in fluid intake beginning a few days before the chemotherapy and continuing for a few days after it.
- If your child is taking oral medication at home, the time of day that medication is given may be critical. Some are best given in the morning, some at midday, some on a full stomach. Be sure to ask your doctor when and how medications should be administered.

Information provided by the
National Cancer Institute
National Institutes of Health

Two Routes to Breast Cancer Detection
Nutrition and Exercise Tips to Reduce Breast Cancer Risk
Reducing Breast Cancer Risks
New Treatment for Early Stage Breast Cancer Reduces Radiation Therapy Duration by 85%
New Surgical Procedure for Breast Cancer Patients
Breast Cancer Treatments Offer Patients Many Options
Prostate Cancer Surgery: Weighing the Risks and Benefits
High Volume Breast Cancer Surgeons Have Better Outcomes
Breast Cancer and Lymph Nodes
Reducing Breast Cancer Risks
Risk Factors for Breast Cancer
The Dangers of Breast Cancer
Travel Distance Might Affect Breast Cancer Treatment
Nutrition During Childhood Cancer
Tamoxifen and Breast Cancer

Figure 9. Example of an article recommendation list

3.6.2 Inferential analysis

In this study, inferential statistical analysis methods were adopted to test the differences between the recommendation lists evaluated and ranked from the experts, and the recommendation lists generated from the proposed recommendation system.

The Kendall (1938) rank correlation analysis method examines the degree of similarity between two sets of rank data given to a same set of elements. The value of rank correlation depends upon the number of inversions of pairs of elements that would be needed to transform one rank order into the other.

Notice that the data on the recommended lists from the recommendation system are ranked data, and the data on the evaluated list from the experts are also ranked data. In addition, the ranked data on both lists are the same elements. In other words, the same elements of the list receive multiple treatments. As a result, the Kendall rank correlation analysis method is used to test the differences between the recommendation lists evaluated and ranked from the experts, and the recommendation lists generated from the proposed recommendation system. The value of Kendell's Tau ranges from -1 to 1. When $\tau b = -1$, it suggests a perfect negative relationship between two variables, whereas $\tau b = -1$ suggests a perfect positive relationship between the two variables. There are several suggestions when interpreting Kendell's Tau. The SPSS tutorial (" Kendall's Tau – Simple Introduction 2021," 2021) suggests when $\tau b > 0.07$, where is a weak positive association; when $\tau b > 0.21$, there is a medium positive association; when $\tau b > 0.35$, there is a strong positive association between 2 variables.

Two outcomes from the value of Kendall's Tau and the p-value were used in the study. In the study, the significant level of the inferential statistical test was 0.05. Therefore, if the p-value was smaller than 0.05, it means that these two lists were significantly correlated. It can be concluded that these two lists were highly consistent. However, the value of Kendall's Tau was also considered. The SPSS tutorial (" Kendall's Tau – Simple Introduction 2021," 2021) suggests when $\tau b > 0.07$, where is a weak positive association; when $\tau b > 0.21$, there is a medium positive association; when $\tau b > 0.35$, there is a strong positive association between 2 variables.

Therefore, when τb was larger than 0.35, it still can be concluded that the two lists are consistent even the p-value was larger than 0.05.The hypotheses to be tested are listed in Chapter 1.3. Table 3 summarizes the relationships among research questions, hypotheses, and related inferential analysis methods.

Research Questions	**Hypotheses posted to answer Research Questions**	**The inferential analysis method used to test hypotheses**
RQ1	H_{01}, H_{02}, and H_{03}	The Kendall rank correlation analysis method
RQ2	H_{04}, H_{05}, and H_{06}	The Kendall rank correlation analysis method
RQ3	H_{07}, H_{08}, and H_{09}	The Kendall rank correlation analysis method

Table 3. The relationships among research questions, hypotheses, and related inferential analysis methods.

Chapter 4. Results

This chapter presents the results of transaction log analysis, Markov chain analysis, and statistical analysis on this research exploration of how health consumers seek information on health information portals. The transaction log analysis is used to extract traversal paths from the log. The Markov chain analysis is employed to model the traversal paths and generate the recommendation lists. Then the statistical analysis is applied to learn whether the proposed recommendation lists can be used to effectively recommend related items to the users.

This chapter is organized as follows: firstly, presentation of the descriptive analysis of the raw data; then discussion of the research questions and the related inferential analyses; finally, presentation of the summary of the results.

4.1 Descriptive analysis of the collected data

As discussed in the data collection section, three out of forty-seven topics were selected to analyze the proposed recommendation system. They are Topic Heart, Topic Infection, and Topic Cancer. Each topic recommendation list includes twenty recommended topics, and five experts judged each recommended topic for relevance. Therefore, 300 relevance scores were collected for the topic level. For the article level, five articles were selected from each topic to evaluate the proposed recommendation system. Each article comes with an article recommendation list. Each list includes thirty recommended articles. And each recommended article was judged by five raters. As a result, 2250 relevance scores were collected for the article level. Similarly, 2250 relevance scores were collected for the Q&A level. Table 4 shows the number of relevance scores collected for each level.

	Topic	**Article**	**Q&A item**

Relevance scores collected	300	2250	2250

Table 4. The collected relevance scores for three research questions

The next section presented the descriptive data analysis for the topic level.

4.1.1 Descriptive data analysis for topic level

The analysis of the topic level includes three topics: Topic Heart, Topic Cancer, and Topic Infection. Each topic has a recommendation topic list. Appendix I contains the recommendation lists for the topic level.

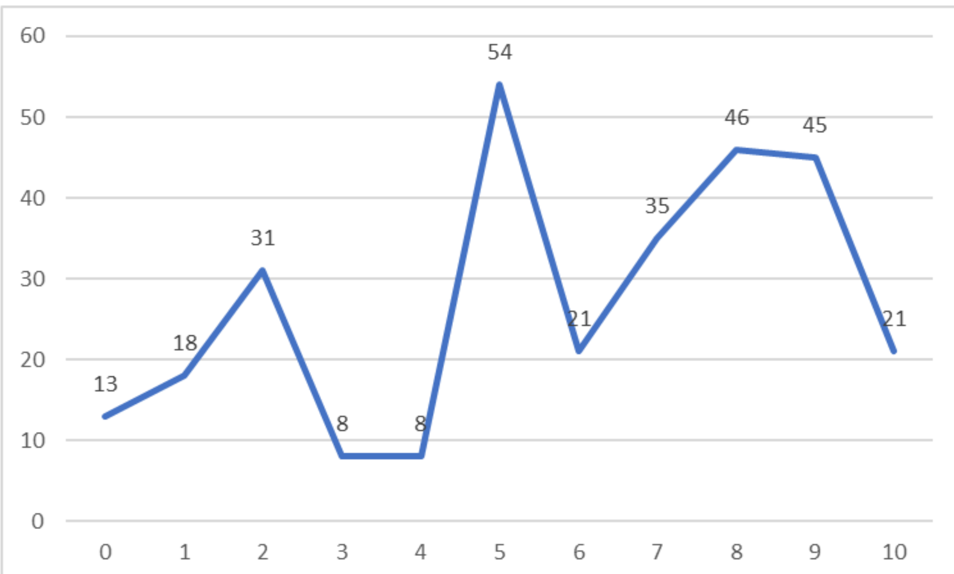


Figure 10. The display of the distribution of the relevance score frequencies for the topic level.

Figure 10 shows the distribution of the relevance score frequencies for the topic level. In Figure 10, the *X*-axis is the relevance score while the *Y*-axis is the frequency of the relevance score. The least often occurring relevance score is three and four, both appeared only eight times. The most frequently occurring score is five, which occurred fifty-four times.

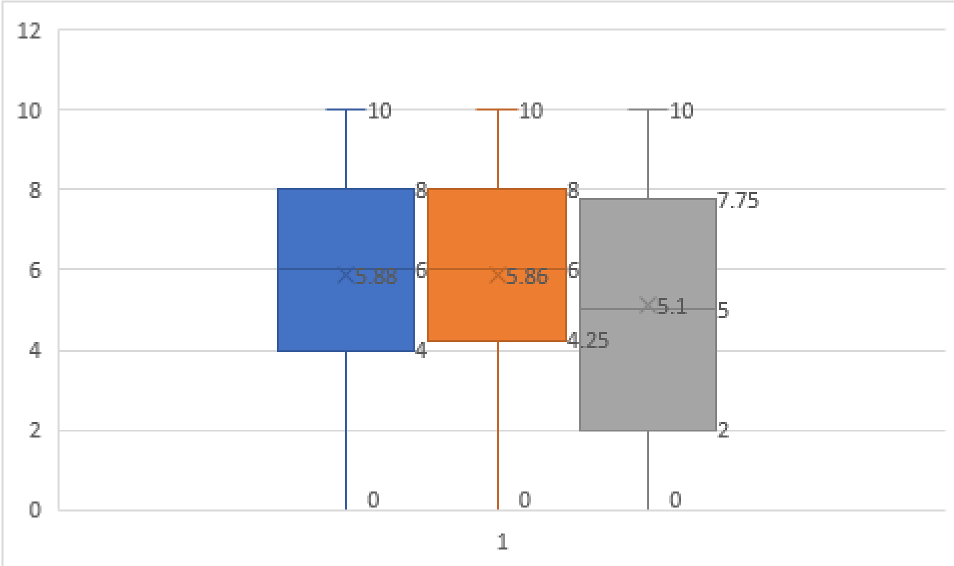


Figure 11. The box and whisker plot for the topic level

Figure 11 shows the box and whisker plot for the topic level. In Figure 11, the *X*-axis is the different topics, whereas the blue plot stands for the Topic Cancer, the orange plot stands for the Topic Heart, and the grey plot stands for the Topic Infection. The *Y*-axis is the relevance score. The mean relevance score is 5.61 for all three topics while the standard deviation is 2.96. The Topic Cancer has the highest mean, 5.88 while the standard deviation is 2.83. The Topic Infection has the lowest mean, which is 5.1 while the standard deviation is 3.06. The mean of Topic Heart is 5.86 while the standard deviation is 2.96. The Topic Cancer and the Topic Heart have the same median relevance score, six. The median score of Topic Infection is five. The maximum score of Topic Cancer is ten, the minimum score is zero. The maximum relevance score of Topic Heart is ten, the minimum score is zero. The maximum relevance score collected from Topic Infection is ten, the minimum score is zero.

4.1.2 Descriptive data analysis for article level

This section presents the descriptive data analysis for the relevance scores collected from article recommendation lists. Five articles were selected from each topic, therefore, there are fifteen articles to be analyzed in total. Table 5 shows the names of the selected articles.

Heart	**Cancer**	**Infection**
American Heart Association Dietary Guidelines	Individual and Marital Adjustment in Married Couples with Breast Cancer	Getting Rid of Hepatitis B in the United States
Heart Failure Treatable with Careful Diagnosis	Breast Lumps and Other Changes	Avian Influenza on the Move
Smoking Before Surgery a Dangerous Decision	Cancer Genetics: Finding and Treating Cancer Before It Occurs	The Facts About HIV Infection and AIDS
Using Diet to Lower Your Blood Pressure	Nutrition and Exercise Tips to Reduce Breast Cancer Risk	Flu Season Begins:
Salt: It's Everywhere	Learning from Cancer Survivors	Smallpox Vaccine Includes a Dose of Risk

Table 5. The names of the selected articles

Each article comes with a recommendation list, and each list comes with fifteen related articles to be recommended. The article recommendation lists are included in Appendix II.

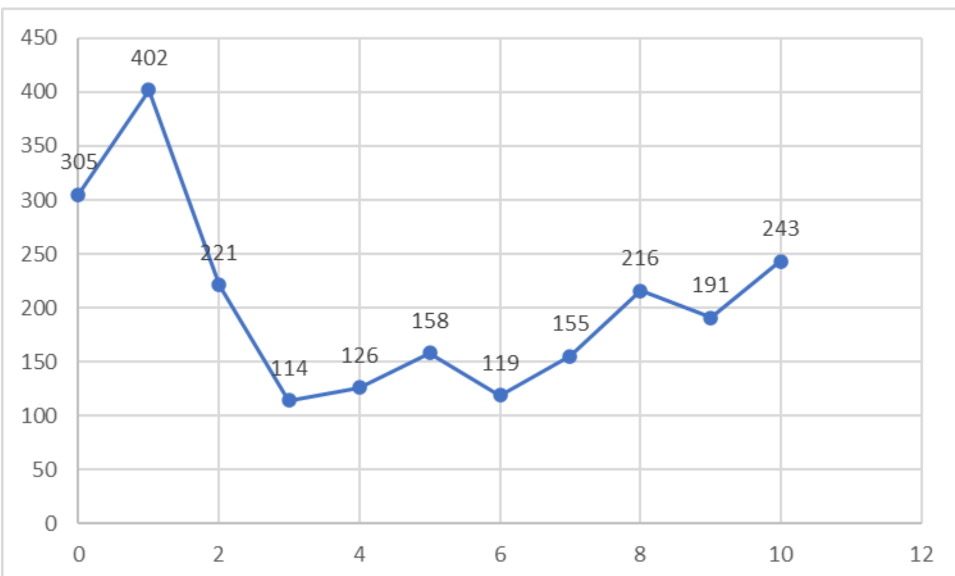


Figure 12, The display of the distribution of the relevance score frequencies for the article level.

Figure 12 illustrates the distribution of the relevance score frequencies for the article level. In Figure 12, the *X*-axis is the relevance score while the *Y*-axis is the frequency of the relevance score. The most often occurred relevance score is 1, which occurred 402 times. The least often occurred relevance score is three, which occurred 114 times.

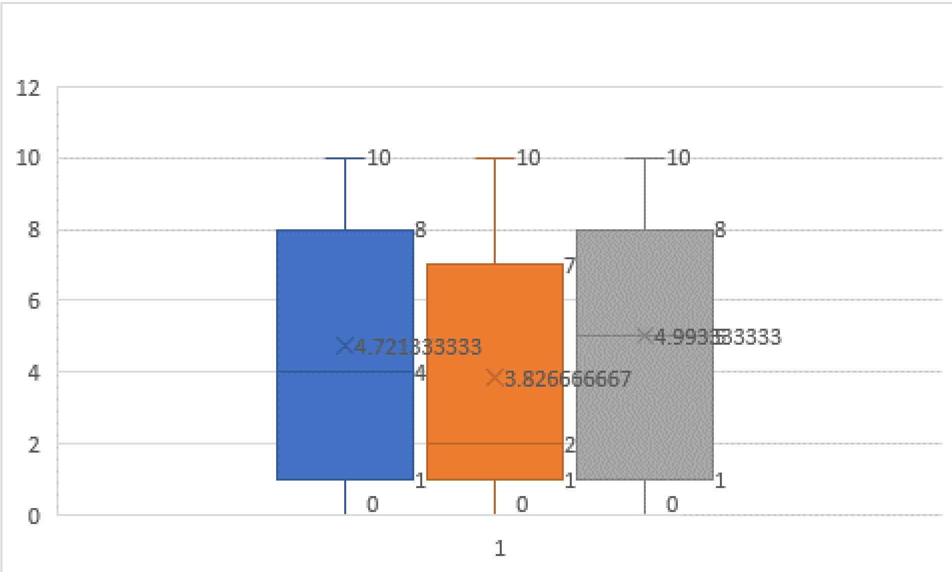


Figure 13. The box and whisker plot for the article level

Figure 13 displays the box and whisker plot for the article level. In Figure 13, the *X*-axis stands for the different topics, whereas the blue plot is the Topic Cancer, the orange plot stands for the Topic Infection, and the grey plot stands for the Topic Heart. The *Y*-axis is the relevance score. The mean relevance score is 4.51 for all three topics while the standard deviation is 3.52. The Topic Heart has the highest mean, 4.99 while the standard deviation is 3.48. The Topic Infection has the lowest mean, which is 3.83 while the standard deviation is 3.49. The Topic Heart also has the largest median score, which is 5. The Topic Infection has the lowest median score, which is two. The median score of Topic Cancer is four. The maximum score of Topic Cancer is ten, the minimum score is zero. The maximum relevance score of Topic Heart is ten, the minimum score is zero. The maximum relevance score collected from Topic Infection is ten, the minimum score is zero.

4.1.3 Descriptive data analysis for Q&A item level

This section presents the descriptive data analysis for the relevance scores collected from Q&A item recommendation lists. Similar to the article recommendation lists, five Q&A items were selected from each topic, therefore, there are fifteen Q&A items to be analyzed. Table 6 shows the names of the selected Q&A items.

Heart	**Cancer**	**Infection**
CPR Not Always the Answer	Breast Cancer and Lymph Nodes	Urinary Tract Infections Common in Women, but Treatable
Cholesterol Particle Test	Calcifications Very Rarely a Sign of Early Breast Cancer	Flu Season Winding Down, But Not Over
Statin Drugs Lower Cholesterol	The Gamma Knife for Non-Invasive Brain Surgery	Bladder Infections More Likely in Women
Is It a Stroke Warning?	Medical Opinions Differ on Testing for Prostate Cancer	Shingles Pain Can Last Long After Rash Heals
Types of Cholesterol	The Dangers of Breast Cancer	Season Has Started, But Flu Shot Still a Good Idea

Table 6. The names of the selected Q&A items

Each Q&A item comes with a recommendation list, and each list comes with fifteen related Q&A items to be recommended. The Q&A item recommendation lists can be found in Appendix III.

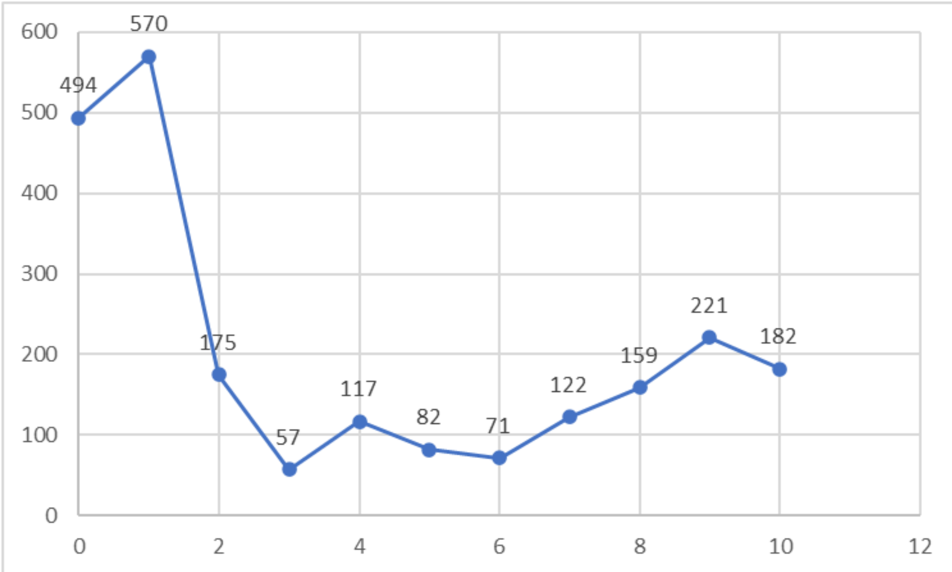


Figure 14. The display of the distribution of the relevance score frequencies for the Q&A items level.

Figure 14 illustrates the distribution of the relevance score frequencies for the Q&A item level. In Figure 14, the *X*-axis is the relevance score while the *Y*-axis is the frequency of the relevance score. The most often occurred relevance score is one, which occurred 570 times. The least often occurred relevance score is three, which occurred 117 times.

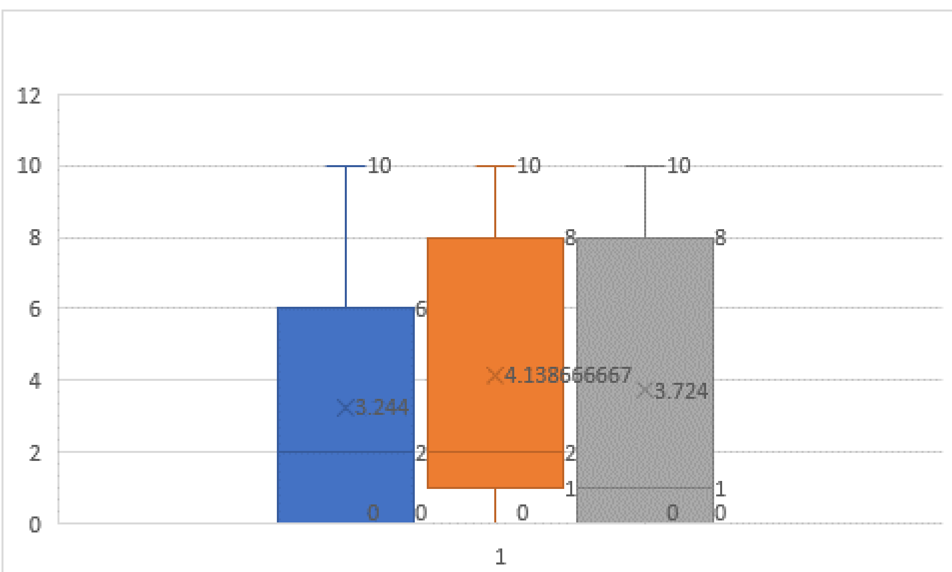


Figure 15. The box and whisker plot for the Q&A item level

Figure 15 displays the box and whisker plot for the Q&A item level. In Figure 15, the *X*-axis stands for the different topics, whereas the blue plot is the Topic Cancer, the orange plot stands for the Topic Infection, and the grey plot stands for the Topic Heart. The *Y*-axis is the relevance score. the *X*-axis is the mean of the relevance score while the *Y*-axis stands for different topics. The mean is 3.70 for all three topics while the standard deviation is 3.63. The Topic Infection has the highest mean, 4.14 while the standard deviation is 3.64. The Topic Cancer has the lowest mean, which is 3.24 while the standard deviation is 3.41. The mean relevance score of Topic Heart is 3.72 while the standard deviation is 3.79. The median relevance scores for the Topic Cancer and the Topic Infection are the same, which is two. The median score of the Topic Heart is one. The maximum score of Topic Cancer is ten, the minimum score is zero. The maximum relevance score of Topic Heart is ten, the minimum score is zero. The maximum relevance score collected from Topic Infection is ten, the minimum score is zero.

4.2 Findings for research question 1

RQ1: Are the topic recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts?

The first research question aims to uncover the hidden traversal patterns between nodes in the subject directory, and to answer whether this hidden traversal pattern uncovered from the HealthLink transaction log can be used to effectively recommend related and relevant topics. To answer this question, the transaction log analysis, the Markov chain analysis, and the inferential analysis were adopted.

4.2.1 Transaction log analysis

In the study, 56,457 sessions with at least two article/Q&A item visiting records were identified. In these sessions, 31,694 sessions contained two article/Q&A item browsing records, which accounted for 56.16% of all sessions; 10,468 sessions contained three article/Q&A item browsing records, which accounts for 18.54%; 4,567 sessions contained four article/Q&A item browsing records, which accounted for 8.09%; 2,744 sessions contain five article/Q&A item browsing records, which accounts for 4.86%; 1,703 sessions contain six article/Q&A item browsing records, which accounts for 3.02%; 1,228 sessions contain seven article/Q&A item browsing records, which accounts for 2.18%; 4,053 sessions contain eight or more article/Q&A item browsing records, which accounts for 7.18% in total.

4.2.2 Markov chain analysis

After cleaning the transaction log data, the session data were imported to the transition matrix. The transition matrix for the topic is a 47*47 matrix since there are forty-seven topics in total. This matrix had 2,209 cells and the total transitions are 621,827 times from topic to topic. The mean of topic transitions in each cell of the matrix was 281.50, the standard deviation was 322.77.

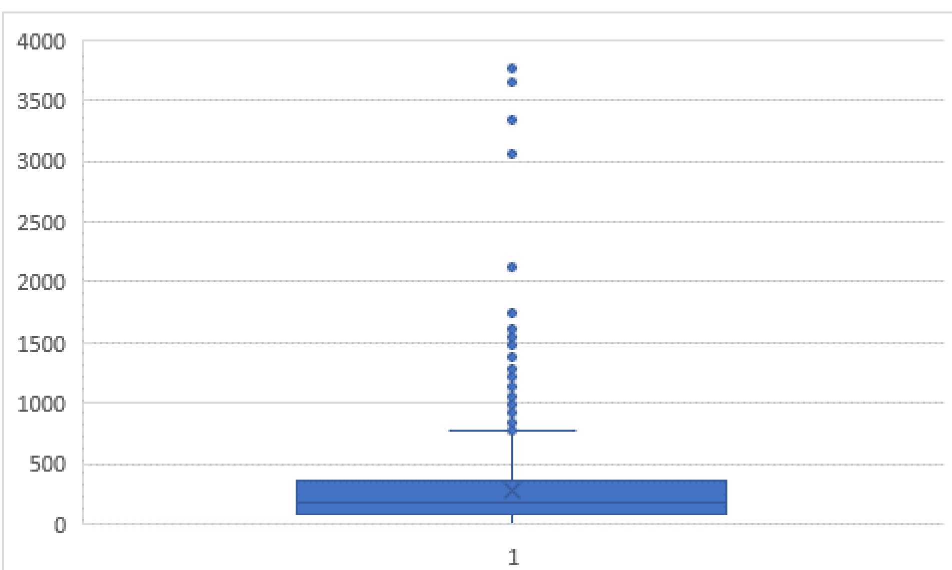


Figure 16. The box and whisker plot for the transitions from topic to topic

Figure 16 shows the box and whisker plot for the transitions from topic to topic. In Figure 16, the *X*-axis stands for the topic transitions, whereas the *Y*-axis is the count of transitions from one topic to another. The median score is 182. The three largest numbers of transitions are, 3775, 3,662, and 3,371. The transitions within *Topic Infection* were 3775; the number of transitions within Topic Heart is 3,662, and the transitions within Topic Cancer is 3371; which are the three topics chosen for analysis in the study. The smallest numbers of transitions are 3, 4, and 4. The number of transitions from Topic Ears/Hearing to Topic Clinical Trials is three. The numbers of transition from *Topic Clinical Trial to Topic Ears/Hearing*, from Topic Immune Disorders to Topic Alternative Medicine, and from Topic Kidney to Topic Clinical Trial are four.

After importing the traversal data into the transition matrix, the modified transition matrix G can be obtained through Equation(20). Then the Markov chain analysis was performed, and the matrix achieved equilibrium after twelve iterations. Once the matrix equilibrium was achieved, a steady vector was obtained. The steady vector then can be ranked. The topic recommendation list consists of the topics that correspond to the top ten largest elements in the vector. The topic recommendation lists were generated for evaluation.

4.2.3 RQ1.1 & H_{01}

It is difficult to evaluate all the topic recommendation lists since there are forty-seven topics in the HealthLink portal. As a result, three topics, *Topic Infection*, *Topic Heart*, and *Topic Cancer,* were selected to answer RQ1. Therefore, three corresponding sub-research questions were proposed.

RQ1.1 asks *"Is the topic recommendation list generated by the proposed recommendation system consistent with the list ranked by the experts in terms of the Infection topic?"*

To answer RQ1.1, the first step was to determine whether the proposed recommendation list is relevant. Therefore, a Wilcoxon signed-rank test was performed to test the differences between the generated recommendation list and a randomly selected topic list. The significance level (α) for the test was equal to 0.05. If the resultant *p*-value of a null hypothesis test was smaller than 0.05, the null hypothesis was rejected. Otherwise, the null hypothesis failed to be rejected. A Wilcoxon signed-rank test indicated that the generated recommendation list was significantly more relevant than the random list $Z=-2.80$, $p<0.005$. The test also indicated that the average relevance score of the generated list, mean=6.96, was significantly higher than the median relevance score of the random list, Mdn= 3.24. Therefore, the null hypothesis was rejected. It suggested that the recommendation list of Topic Infection that generated from the proposed recommendation system was more relevant than the list generated randomly.

Next, Kendall's τ coefficient then was adopted to compare the recommendation list ranked by the proposed recommendation system and the recommendation list ranked by the experts. The SPSS tutorial suggests when $\tau b > 0.07$, there is a weak positive association; when $\tau b > 0.21$, there is a medium positive association; when $\tau b > 0.35$, there is a strong positive association between two variables. Five experts were recruited to rank the recommendation list for Topic Infection. To ensure the reliability of the study, the intraclass correlation coefficient (ICC) reliability formula was adopted to test the inter-coder reliability of five experts. The ICC score for Topic Infection is 0.532, which is a fair agreement. The reliability of the relevant judgment data can be ensured.

The resultant *p*-value of the correlation test is 0.089, which is larger than the significance level (α=0.05). It indicated that the null hypothesis H_{01} failed to be rejected and there is no significant correlation between the two lists. However, Kendall's τ coefficient equals 0.422, which is larger than 0.35, suggesting although H_{01} failed to be rejected, there was still a strong association between the recommendation list ranked by the proposed recommendation system and the list ranked by the experts. In other words, the recommendation list of Topic Infection generated from the proposed recommendation system is consistent with the list ranked by experts.

4.2.4 RQ1.2 & H_{02}

RQ1.2 asks *"Is the topic recommendation list generated by the proposed recommendation system consistent with the list ranked by the experts in terms of the Heart topic?"*

The intraclass correlation coefficient (ICC) reliability formula is adopted to test the inter-coder reliability of five experts. The ICC score is 0.563, which is a fair agreement.

A Wilcoxon signed-rank test was performed to test the differences between the generated recommendation list and the random list. The significance level (α) for the test was equal to 0.05. If the resultant *p*-value of a null hypothesis test was smaller than 0.05, the null hypothesis was rejected. Otherwise, the null hypothesis failed to be rejected. A Wilcoxon signed-rank test indicates that the generated recommendation list was significantly more relevant than the random list $Z=-2.67$, $p<0.008$. The test also indicated that the average relevance score of the generated list, mean=7.7, was significantly higher than the mean relevance score of the random list, mean= 4.0. Therefore, the null hypothesis of the Wilcoxon signed-rank test was rejected. It suggested that the recommendation list of Topic Heart that generated from the proposed recommendation system was more relevant than the list generated randomly.

The ICC score for Topic Heart is 0.563, which is a fair agreement. The reliability of the relevant judgment data can be ensured. The resultant *p*-value of the correlation test is 0.06, which is larger than the significance level (α), suggesting there is no significant correlation between the two lists. As a result, H_{02} failed to be rejected. However, Kendall's τ coefficient equals 0.467, which is larger than 0.35, suggesting a strong association between the list ranked by the proposed recommendation system and the list ranked by experts. In other words, it suggested a positive correlation between the two lists even though the correlation is not significant. In summary of H_{02}, the recommendation list of Topic *Heart* generated from the proposed recommendation system is consistent with the list ranked by the experts.

4.2.5 RQ1.3 & H_{03}

RQ1.3 asks *"Is the topic recommendation list generated by the proposed recommendation system consistent with the list ranked by the experts in terms of the Cancer topic?"*

A Wilcoxon signed-rank test was performed to test the differences between the generated recommendation list and the random list. The significance level (α) for the test was equal to 0.05. If the resultant *p*-value of a null hypothesis test was smaller than 0.05, the null hypothesis was rejected. Otherwise, the null hypothesis failed to be rejected. A Wilcoxon signed-rank test indicates that the generated recommendation list was not significantly more relevant than the random list at Z=-1.33, p<0.19. Therefore, the null hypothesis failed to be rejected. It suggested that there is no significant difference between the two lists in terms of relevance judgment.

The intraclass correlation coefficient (ICC) reliability formula was adopted to test the inter-coder reliability of five experts. The ICC score for Topic Infection is 0.449, which is a fair agreement. The reliability of the relevant judgment data can be ensured. Kendall's τ coefficient equals 0.6, which is larger than 0.35, suggesting a strong association between the list ranked by the proposed recommendation system and the recommendation list ranked by the experts. The resultant *p*-value of the correlation test is 0.016, which is smaller than the significance level (α), suggesting there is significant correlation between the two lists. In other words, it suggested that there is a positive correlation between the two lists.

The results of H_{03} are more complicated to interpret since the null hypothesis of the Wilcoxon signed-rank failed to be rejected. However, the mean relevance score (6.62) of the recommendation list generated from the proposed system was still larger than the mean relevance score (5.14) of the randomly generated list. And more important, there is a significant positive correlation between the list ranked by the proposed system and the list ranked by experts. As a result, H_{03} was rejected, and the recommendation list of Topic Cancer generated from the proposed recommendation system is highly consistent with the list ranked by the experts.

4.2.6 Summary

The first research question aimed to answer whether the recommendation lists of topics generated from the proposed system are consistent with the lists ranked by experts. To answer this question, five experts were recruited to rank the recommendation lists. The relevance of the recommendation lists generated from the proposed system was confirmed by conducting the

Wilcoxon signed-rank tests. Then the lists ranked by the proposed system and the lists ranked by experts are compared. If there is a positive correlation between the lists generated from the system and the lists ranked by experts, it can be concluded that the topic recommendation lists generated from the proposed system are consistent with the lists ranked by experts.

Since there are forty-seven topics on the portal. Three topics, Topic Infection, Topic Heart, and Topic Caner were sampled to represent all topics. The first research question then was divided into three sub-questions, each sub-question representing one topic.

Table 7 shows the summary of the statistical results for the Wilcoxon signed-rank tests of RQ1. It shows that ICC for all three topics were fair agreement. Therefore, the reliability of the study can be ensured. The results of Wilcoxon signed-rank tests of Topic Infection and Topic Heart were rejected. They suggested that the topic recommendation lists generated from the proposed system are more relevant than the recommendation lists generated randomly in terms of the Topic Infection and Topic Heart. However, the results of the Wilcoxon signed-rank test of Topic Cancer failed to be rejected and the recommendation list generated from the proposed system is not significantly more relevant than the list generated randomly.

Topic	ICC	Test	p-value	Z
Infection	0.532	Wilcoxon signed-rank test	0.005	-2.80
Heart	0.563	Wilcoxon signed-rank test	0.008	-2.67
Cancer	0.449	Wilcoxon signed-rank test	0.19	-1.33

Table 7. The statistical results for the Wilcoxon signed-rank tests of RQ1

Table 8 shows the summary of findings for RQ1. The null hypotheses of H_{01} and H_{02} failed to be rejected since both *p*-values were larger than the significance value (α=0.05). However, the Kendall's τ coefficients of Topic Infection and Topic Heart are 0.422 and 0.467 respectively, both were higher than 0.35. It suggested there is a strong positive correlation between the recommendation lists ranked by the proposed system and the recommendation lists ranked by experts. It was concluded that the topic recommendation lists generated from the proposed system are consistent with the lists ranked by the experts in terms of Topic Infection and Topic *Heart*. For H_{03}, although the result of the Wilcoxon signed-rank test of Topic Cancer failed to be rejected, the mean relevance score of the recommendation list generated from the proposed system is larger than the recommendation list generated randomly. More importantly, H_{03} was rejected since the p-value was smaller than the significance level (α=0.05). And the Kendall's τ coefficient of Topic cancer is 0.6, which is larger than 0.35. It suggested that the topic recommendation list generated from the proposed system is highly consistent with the list ranked by experts in terms of Topic Cancer.

Questions	**Hypothesis**	**Test**	**p-value**	**Kendall's τ**	**Results**	**Consistency with experts**
RQ1.1	H_{01}	Correlation test	0.089	0.422	Not rejected	Consistent
RQ1.2	H_{02}	Correlation test	0.06	0.467	Not rejected	Consistent
RQ1.3	H_{03}	Correlation test	0.016	0.6	Rejected	Highly consistent

Table 8. Summary of the findings for RQ 1

As a result, it can be concluded that the topic recommendation lists generated from the proposed system are consistent with the lists ranked by the experts.

4.3 Findings for research question 2

RQ2: Are the article recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts?

The second research question aims to unveil the hidden traversal patterns between articles on the health portal, and whether these hidden patterns can be used to effectively recommend related and relevant articles. Similar to RQ1, the transaction log analysis, the Markov chain analysis, and the inferential analysis were adopted.

4.3.1 Transaction log analysis and Markov chain analysis

Numerous sessions were recorded in the transaction log. However, only sessions with at least two articles visiting records were used for analysis since the creation of a transition matrix requires at least one jump from one article to another. In the study, 56,457 sessions with at least two article visiting records were identified from the transaction log. The session data were imported to the transition matrices. Fifteen articles were selected from three topics, Topic Infection, Topic Heart, and Topic Cancer for analysis. Each article came with one transition matrix. Therefore, there were fifteen transition matrices for article-level analysis. Although there are 56,457 sessions identified from the transaction log. The article-article transition matrices were still sparse. Therefore, the modified transition matrix G can be obtained through Equation (20) to solve this problem. Then the Markov chain analysis was performed. When the matrix equilibrium was achieved, a steady vector was obtained. The elements of the steady vector then can be ranked. The top fifteen largest elements in the vector were obtained, each corresponding to an article. The article recommendation list consisted of these fifteen articles in a descending order in terms of the value of the corresponding element. The article recommendation lists were used for evaluation.

4.3.2 RQ2.1 & H_{04}

RQ2.1 asks *"Are the article recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts in terms of the Infection topic?"*

To answer RQ2.1, it is crucial to make sure the proposed article recommendation lists are relevant. Five articles were randomly selected from Topic Infection to represent this topic. Five Wilcoxon signed-rank tests were performed to test the differences between the generated recommendation lists and the random lists. The significance level (α) for the test was equal to 0.05. If the resultant *p*-value of a null hypothesis test was smaller than 0.05, the null hypothesis was rejected. Otherwise, the null hypothesis failed to be rejected. Table 9 displays the statistical results of the Wilcoxon signed-rank tests. As a result, all five null hypotheses were rejected. It suggests there were significant differences between the recommendation lists generated by the proposed recommendation system and the randomly generated lists. In other words, the recommendation lists generated by the proposed recommendation system were more relevant than the lists generated randomly.

The intraclass correlation coefficient (ICC) reliability formula was adopted to test the inter-coder reliability of five experts. The ICC values of the five articles are 0.606, 0.781, 0.702, 0.737, and 0.764 respectively, which are larger than 0.6, indicating a good or an excellent agreement among the experts.

Article	**ICC**	**Test**	**p-value**	**Z**
A1	0.606	Wilcoxon signed-rank test	0.001	-3.409

A2	0.781	Wilcoxon signed-rank test	0.001	-3.414
A3	0.702	Wilcoxon signed-rank test	0.001	-3.238
A4	0.737	Wilcoxon signed-rank test	0.001	-3.412
A5	0.764	Wilcoxon signed-rank test	0.001	-3.413

Table 9. The statistical results for the Wilcoxon signed-rank tests of RQ2.1

Kendall's τ coefficient then was adopted to compare the recommendation lists ranked by the proposed recommendation system and the recommendation lists ranked by the experts. The results of the correlation test are shown in Table 10. The *p*-values of five correlation tests are smaller than the significance level (α), suggesting there were significant correlations between the two lists. Kendall's τ coefficients were 0.695, 0.562, 0.638, 0.524, and 0.562 respectively, all are larger than 0.35, suggesting a strong positive association between the recommendation lists ranked by the proposed system and ranked by the experts.

Combining the results from Wilcoxon signed-rank test and Kendall's τ coefficients, H_{04} was rejected. There are significant correlations between the article lists ranked by the experts and lists generated by the proposed recommendation system in terms of Topic Infection. In other words, the recommendation lists of Topic Infection generated from the proposed system are highly consistent with the lists ranked by the experts.

Article	**Test**	**p-value**	**Kendall's τ**	**Consistency with experts**
A1	Correlation test	0.000	0.695	Highly consistent
A2	Correlation test	0.004	0.562	Highly consistent
A3	Correlation test	0.001	0.638	Highly consistent
A4	Correlation test	0.006	0.524	Highly consistent
A5	Correlation test	0.004	0.562	Highly consistent

Table 10. The correlation results for RQ2.1

4.3.3 RQ2.2 & H_{05}

RQ 2.2 asks "*Are the article recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts in terms of the Heart topic?*"

To address RQ2.2, the first step is to ensure the recommendation lists are relevant. Five Wilcoxon signed-rank tests were conducted on the corresponding articles selected from the topic. The Wilcoxon signed-rank tests were performed to test the differences between the generated recommendation lists and the random lists. The significance level (α) for the test was equal to 0.05. If the resultant *p*-value of a null hypothesis test was smaller than 0.05, the null hypothesis was rejected. Otherwise, the null hypothesis failed to be rejected. Table 11 displays the results of the inferential statistical tests. It suggested the five corresponding null hypotheses were rejected, there were significant differences between the recommendation lists generated by the proposed recommendation system and the randomly generated lists. The recommendation lists generated by the proposed system are relevant. The intraclass correlation coefficient (ICC) reliability formula was adopted to test the inter-coder reliability of five experts. The ICC values of the five articles are 0.502, 0.403, 0.883, 0.71, and 0.818 respectively. The ICC values of articles A3, A4, and are larger than 0.6, indicating a good agreement among experts. The ICC values of articles A1 and A2 are larger than 0.4 while smaller than 0.6, indicating a fair agreement among experts. It means the relevance scores collected from experts are reliable to conduct the study.

Article	ICC	Test	p-value	Z

A6	0.502	Wilcoxon signed-rank test	0.001	-3.352
A7	0.403	Wilcoxon signed-rank test	0.001	-3.354
A8	0.883	Wilcoxon signed-rank test	0.001	-3.412
A9	0.71	Wilcoxon signed-rank test	0.001	-3.417
A10	0.818	Wilcoxon signed-rank test	0.001	-3.412

Table 11. The statistical results for the Wilcoxon signed-rank tests of RQ2.2

Kendall's τ was used to compare the recommendation lists ranked by experts and by the proposed system. As shown in table 12, Kendall's τ is 0.429. 0.619. 0.81, 0.276, and 0.524 for five selected articles, four of them are larger than 0.35, indicating a strong positive association between the list ranked by the recommendation system and the recommendation list ranked by experts. The *p*-values of these four correlation tests are smaller than the significance level (α=0.05), suggesting there are significant correlations between the two lists. Therefore, the null hypotheses of A6, A7, A8, and A10 were rejected, there are significant correlations between the article recommendation lists generated from the proposed system and the lists ranked by the experts in terms of these four articles. In other words, the article recommendation lists generated from the proposed system are highly consistent with the lists ranked by the experts in terms of articles A6, A7, A8, and A10. The *p*-values of A9's correlation tests are larger than the significance level (α=0.05), suggesting there are no significant correlations between the two lists. The Kendall's τ of A9 was larger than 0.21, which suggests a medium positive association. Therefore, Although the null hypothesis of the article A9 failed to be rejected, the article recommendation list generated from the proposed system is still consistent with the list ranked by the experts in terms of the article A9.

For Topic Heart, the null hypotheses of the correlation tests for the four articles (article A6, A7, A8, and A10) were rejected, the null hypothesis of the correlation test for the one article

(article A9) failed to be rejected. As a result, H_{05} failed to be rejected. There are no significant correlations between the recommendation lists generated from the proposed system and the lists ranked by the experts in terms of Topic Heart. However, the results of Kendall's τ still suggested there are high or medium association between the two lists. As a result, the article recommendation lists generated by the proposed recommendation system are consistent with the lists ranked by the experts in terms of the Heart topic.

Article	**Test**	**p-value**	**Kendall's τ**	**Consistency with experts**
A6	Correlation test	0.026	0.429	Highly consistent
A7	Correlation test	0.002	0.619	Highly consistent
A8	Correlation test	0.000	0.81	Highly consistent
A9	Correlation test	0.151	0.276	Consistent
A10	Correlation test	0.006	0.524	Highly consistent

Table 12. The correlation results for RQ2.2

4.3.4 RQ2.3 & H_{06}

RQ 2.3 asks "*Are the article recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts in terms of the Cancer topic?*"

To answer RQ 2.3, it's crucial to ensure the article recommendation lists are relevant. Five Wilcoxon signed-rank tests were conducted on the corresponding articles selected from the topic. The Wilcoxon signed-rank tests were performed to evaluate the differences between the generated recommendation lists and the random lists. The significance level (α) for the test was equal to 0.05. If the resultant *p*-value of a null hypothesis test was smaller than 0.05, the null hypothesis was rejected. Otherwise, the null hypothesis failed to be rejected. Table 13 summarized the results of Wilcoxon signed-tank tests. It indicated that all five null hypotheses were rejected, and there were significant differences between the recommendation lists generated by the proposed recommendation system and the randomly generated lists. The recommendation lists generated by the proposed system are relevant. The intraclass correlation coefficient (ICC) reliability formula was adopted to test the inter-coder reliability of five experts. The ICC values of the five articles are 0.633, 0.864, 0.652, 0.561, and 0.644 respectively, indicating a fair, or a good or an excellent agreement among experts. It means the relevance scores collected from the experts are reliable to conduct the study.

Article	**ICC**	**Test**	**p-value**	**Z**
A11	0.633	Wilcoxon signed-rank test	0.004	-2.869
A12	0.864	Wilcoxon signed-rank test	0.001	-3.408
A13	0.652	Wilcoxon signed-rank test	0.002	-3.154
A14	0.561	Wilcoxon signed-rank test	0.006	-2.727
A15	0.604	Wilcoxon signed-rank test	0.001	-3.411

Table 13. The statistical results for the Wilcoxon signed-rank tests of RQ2.3

Kendall's τ was used to compare the recommendation lists ranked by experts and by the proposed system. As shown in table 14, The *p*-values of these five correlation tests are smaller than the significance level (α=0.05), suggesting there are significant correlations between the two lists. Kendall's τ is 0.695, 0.41, 0.733, 0.724, and 0.448 for five selected articles, all of which are larger than 0.35, indicating a strong positive association between the list ranked by the recommendation system and the recommendation list ranked by experts.

As a result, H_{06} was rejected. There are significant correlations between the article recommendation lists generated from the proposed system and the lists ranked by the experts in terms of the Cancer topic. The article recommendation lists generated from the proposed system are consistent with the lists ranked by the experts in terms of the Cancer topic.

Article	**Test**	**p-value**	**Kendall's τ**	**Consistency with experts**
A11	Correlation test	0.000	0.695	Highly consistent
A12	Correlation test	0.033	0.41	Highly consistent
A13	Correlation test	0.000	0.733	Highly consistent
A14	Correlation test	0.000	0.724	Highly consistent
A15	Correlation test	0.020	0.448	Highly consistent

Table 14. The correlation results for RQ2.3

4.3.5 Summary

The second research question aimed to answer whether the recommendation lists of articles generated from the proposed system are consistent with the lists ranked by experts. To answer this question, five experts were recruited to rank the recommendation lists. The Wilcoxon signed-rank tests were used to confirm the relevance of the recommendation lists generated from the proposed system. Then the lists ranked by the proposed system and the lists ranked by experts were compared. If there is a positive correlation between the two lists, it can be concluded that the topic recommendation lists generated from the proposed system are consistent with the lists ranked by experts.

Since the article on the portal were classified into forty-seven topics. Articles from the three topics (Topic Infection, Topic Heart, and Topic Caner) were sampled to represent all the forty-seven topics. The second research question then was divided into three sub-questions, each sub-question representing one topic.

The results of the tests can be found in Tables 9-14. The ICC method was used to ensure the relevance scores data collected from five experts are reliable. According to Cicchetti (1994), when the ICC score is larger than 0.75, it suggests an excellent inter-rater agreement; when the ICC score is between 0.6 and 0.75, it indicates a good inter-rater agreement; when the ICC score is between 0.4 and 0.59, it indicates a fair inter-rater agreement; when the ICC score is less than 0.4, it suggests a poor inter-rater agreement. In the study, the ICC scores for five articles from Topic Infection are larger than 0.6, which suggests an excellent or a good agreement among these five experts. The ICC scores for A8-A10 from Topic Heart are larger than 0.6, indicating an excellent or a good agreement. The ICC scores for A6 and A& from Topic Heart are between 0.4 and 0.59. indicating a fair agreement. Therefore, the inter-rater agreement for RQ2 is either excellent, good, or fair. It concludes that the relevant scores data collected for RQ2 are reliable.

All the null hypotheses of Wilcoxon signed-rank tests were rejected. They suggested that the article recommendation lists generated from the proposed system are more relevant than the recommendation lists generated randomly.

Table 15 summarized the findings for RQ2. The null hypothesis of the correlation test for the article A9 failed to be rejected. The other null hypotheses of the correlation tests for the other fourteen articles were rejected. The Kendall's τ coefficient of A9 from Topic Heart is 0.276, which is larger than 0.21 but smaller than 0.35, indicating a medium positive association between the recommendation list ranked by experts and the list ranked by the proposed system.

The Kendall's τ coefficients for the other fourteen articles are larger than 0.35, indicating a strong association between the recommendation list ranked by experts and the list ranked by the proposed system.

As a result, it can be concluded the article recommendation lists generated from the proposed system are consistent with the lists ranked by the experts.

Questions	**Hypothesis**	**Test**	**Results**
RQ2.1	H_{04}	Correlation test	A1: Rejected
			A2: Rejected
			A3: Rejected
			A4: Rejected
			A5: Rejected
RQ2.2	H_{05}	Correlation test	A6: Rejected
			A7: Rejected
			A8: Rejected
			A9: Not Rejected
			A10: Rejected
RQ2.3	H_{06}	Correlation test	A11: Rejected
			A12: Rejected
			A13: Rejected
			A14: Rejected
			A15: Rejected

Table 15. Summary of findings for RQ2

4.4 Findings for research question 3

RQ3: Are the Q&A item recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts?

The last research question was proposed to undercover the hidden traversal patterns between Q&A items on the health portal, and whether these hidden patterns can be used to effectively recommend related and relevant Q&A items. Similar to RQ1 and RQ2, the transaction log analysis, the Markov chain analysis, and the inferential analysis were adopted.

4.4.1 Transaction log analysis and Markov chain analysis

Like RQ 2, 56,457 sessions with at least two Q&A item visiting records were identified from the transaction log. The session data were imported to the corresponding transition matrices. Fifteen Q&A items were selected from three topics, Topic Infection, Topic Heart, and Topic Cancer for analysis. Each Q&A item came with one transition matrix. Therefore, there are fifteen transition matrices for Q&A items -level analysis. Although there are 56,457 sessions identified from the transaction log. The Q&A item - Q&A item transition matrices are still sparse. Therefore, the modified transition matrix G can be obtained through Equation (20) to solve this problem. Then the Markov chain analysis was performed. When the matrix equilibrium was achieved, a steady vector was obtained. The elements of the steady vector then can be ranked. The Q&A items recommendation list consisted of the Q&A items that correspond to the top fifteen largest elements in the vector. The Q&A item recommendation lists were generated for evaluation.

4.4.2 RQ3.1 & H_{07}

RQ3.1 asks "*Are the Q&A item recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts in terms of the Infection topic?*"

To answer RQ3.1, the recommendation lists are evaluated and ranked by five experts. 750 relevance scores were collected for this analysis. The intraclass correlation coefficient ICC reliability formula is used to test the agreement among experts. The ICC values of five Q&A items are 0.854, 0.81, 0.816, 0.926, and 0.756 respectively. All are larger than 0.6, indicating a strong agreement among these five experts. Then five Wilcoxon signed-rank tests were conducted to investigate the differences between the recommendation lists generated by the proposed system and the recommendation lists generated randomly. As shown in Table 16, all five null hypotheses were rejected, there were significant differences between the recommendation lists generated by the proposed recommendation system and the randomly generated lists. Therefore, the recommendation lists generated by the proposed system are relevant.

Q&A item	**ICC**	**Test**	**p-value**	**Z**
Q1	0.854	Wilcoxon signed-rank test	0.001	-3.418
Q2	0.81	Wilcoxon signed-rank test	0.001	-3.357
Q3	0.816	Wilcoxon signed-rank test	0.001	-3.238
Q4	0.926	Wilcoxon signed-rank test	0.001	-3.415
Q5	0.756	Wilcoxon signed-rank test	0.001	-3.408

Table 16. The statistical results for the Wilcoxon signed-rank tests of RQ3.1

In Table 17, the *p*-values of these five correlation tests are smaller than the significance level (α=0.05), suggesting there are significant correlations between the two lists. Kendall's τ is 0.41, 0.638, 0.829, 0.562, and 0.79 for five Q&A items. All of them are larger than 0.35, indicating a strong positive correlation between the recommendation lists ranked by experts and the lists ranked by the proposed system.

As a result, H_{07} was rejected. There are significant correlations between the Q&A item recommendation lists ranked by the experts and the lists generated by the proposed recommendation in terms of the Infection topic. In other words, the Q&A recommendation lists from the Infection topic generated from the proposed recommendation system are consistent with the lists ranked by the experts.

Article	**Test**	**p-value**	**Kendall's τ**	**Consistency with experts**
Q1	Correlation test	0.033	0.41	Highly Consistent
Q2	Correlation test	0.001	0.638	Highly Consistent
Q3	Correlation test	0.000	0.829	Highly Consistent
Q4	Correlation test	0.004	0.562	Highly Consistent
Q5	Correlation test	0.000	0.79	Highly Consistent

Table 17. The correlation results for RQ3.1

4.4.3 RQ3.2 & H_{08}

RQ3.2 asks "*Are the Q&A item recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts in terms of the Heart topic?*"

The intraclass correlation coefficient (ICC) reliability formula was adopted to assess the inter-coder reliability of five experts. The ICC scores for five Q&A items from Topic Heart are 0.845, 0.883, 0.875, 0.866, and 0.832, all of which are larger than 0.6, indicating a strong agreement among experts. The relevance score data collected for RQ3.2 is reliable. The results of Wilcoxon signed-rank tests are shown in Table 18. It suggests all five null hypotheses were rejected, there were significant differences between the recommendation lists generated by the proposed recommendation system and the randomly generated lists. Therefore, the recommendation lists generated by the proposed system are relevant.

Q&A item	ICC	Test	p-value	Z
Q6	0.845	Wilcoxon signed-rank test	0.001	-3.411
Q7	0.883	Wilcoxon signed-rank test	0.001	-3.415
Q8	0.875	Wilcoxon signed-rank test	0.001	-3.300
Q9	0.866	Wilcoxon signed-rank test	0.001	-3.411
Q10	0.832	Wilcoxon signed-rank test	0.001	-3.412

Table 18. The statistical results for the Wilcoxon signed-rank tests of RQ3.2

Table 19 displays the results of the correlation tests. The *p*-values of correlation tests for Q6, Q7, Q8, and Q10 are smaller than the significance level (α=0.05), suggesting there are significant correlations between the two lists. The *p*-value of the correlation test for Q9 is larger than the significance level (α=0.05), suggesting there is no significant correlation between the two lists. The results of Kendall's τ of Q&A items from Topic Heart are 0.733, 0.505, 0.696, 0.371, and 0.467, all of which are larger than 0.35, indicating a strong positive correlation between the recommendation lists ranked by experts and the recommendation lists ranked by the proposed system. Therefore, for Q&A items Q6, Q&, Q8, and Q10, the Q&A item recommendation lists generated from the proposed system are highly consistent with the lists ranked by the experts. For Q9, although there is no significant correlation between the two lists, the Kendall's τ value is larger than 0.35, which indicates a strong association. It suggested that the Q&A item recommendation list generated from the proposed system is consistent with the list ranked by the experts in terms of the Q&A item Q9.

As a result, H_{08} failed to be rejected. There are no significant correlations between the Q&A item recommendation lists ranked by the experts and the lists generated by the proposed recommendation system in terms of the Heart topic. However, combining the results of the Wilcoxon signed-rank tests and the results of Kendall's τ, it can be concluded that the Q&A item recommendation lists generated by the proposed recommendation system are consistent with the lists ranked by the experts in terms of the Heart topic.

Article	Test	p-value	Kendall's τ	Consistency with experts
Q6	Correlation test	0.000	0.733	Highly consistent
Q7	Correlation test	0.009	0.505	Highly consistent
Q8	Correlation test	0.000	0.695	Highly consistent
Q9	Correlation test	0.054	0.371	Consistent
Q10	Correlation test	0.015	0.467	Highly consistent

Table 19. The correlation results for RQ3.2

4.4.4 RQ3.3 & H_{09}

RQ3.3 asks "*Are the Q&A item recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts in terms of the Cancer topic?*"

The intraclass correlation coefficient (ICC) reliability formula was adopted to evaluate the inter-coder reliability of five experts. The ICC values of the five articles are 0.72, 0.587, 0.573, 0.636, and 0.891 respectively, all are larger than 0.6, indicating a good agreement among experts. It means the relevance scores collected from experts are reliable to conduct the study. The Wilcoxon signed-rank tests were performed to test the differences between the generated recommendation lists and the random lists. The significance level (α) for the test was equal to 0.05. If the resultant *p*-value of a null hypothesis test was smaller than 0.05, the null hypothesis was rejected. Otherwise, the null hypothesis failed to be rejected. Table 20 summarized the results of Wilcoxon signed-tank tests. It indicated that the null hypotheses for Q11, Q12, Q13, and Q15 were rejected, there were significant differences between the recommendation lists generated by the proposed recommendation system and the randomly generated lists. However, the null hypothesis for Q14 failed to be rejected. It suggests that there are no significant differences between the recommendation list generated from the proposed system and the recommendation list generated randomly for Q14. In other words, the list generated from the proposed system is not significantly more relevant than the list randomly generated in terms of the Q&A item Q14.

Q&A item	**ICC**	**Test**	**p-value**	**Z**
Q11	0.72	Wilcoxon signed-rank test	0.001	-3.414
Q12	0.587	Wilcoxon signed-rank test	0.002	-3.068
Q13	0.573	Wilcoxon signed-rank test	0.001	-3.298
Q14	0.636	Wilcoxon signed-rank test	0.054	-1.749

Q15	0.891	Wilcoxon signed-rank test	0.001	-3.354

Table 20. The statistical results for the Wilcoxon signed-rank tests of RQ3.3

Table 21 displays the results of correlation tests and Kendall's τ. Kendall's τ was used to compare the recommendation lists ranked by experts and by the proposed system. As shown in table 21, Kendall's τ is 0.695, 0371, 0.619, 0.371, and 0.638 for five selected articles, all of which are larger than 0.35, indicating a strong positive association between the list ranked by the recommendation system and the recommendation list ranked by experts. The *p*-values of correlation tests for Q11, Q13, and Q15 are smaller than the significance level (α), suggesting there are significant correlations between the two lists. In other words, the lists generated from the proposed system were consistent with the lists ranked by experts in terms of Q&A items Q11, Q13 and Q15. The *p*-value of correlation tests for Q12 and Q14 are larger than the significance level (α), suggesting there are no significant correlations between the two lists. However, considering the Kendall's τ for Q12 and Q14 are larger than 0.35, it can be concluded that the lists generated from the proposed system were consistent with the lists ranked by experts in terms of Q&A items Q12 and Q14.

Article	**Test**	**p-value**	**Kendall's τ**	**Consistency with experts**
Q11	Correlation test	0.000	0.695	Highly consistent
Q12	Correlation test	0.054	0.371	Consistent
Q13	Correlation test	0.001	0.619	Highly consistent
Q14	Correlation test	0.054	0.371	Consistent
Q15	Correlation test	0.001	0.638	Highly consistent

Table 21. The correlation results for RQ3.3

Although the null hypothesis of the Wilcoxon signed-rank test for Q14 failed to be rejected, the mean of the relevance score of the recommendation list generated from the proposed system (2.99) is larger than the mean of the recommendation list generated randomly (1.25). It indicated that the recommendation list generated from the proposed system is more relevant than the recommendation list generated randomly, although not statistically significant.

Combing the results of the Wilcoxon signed-rank tests and the correlation tests, H_{09} failed to be rejected. There are no significant correlations between the Q&A item recommendation lists ranked by the experts and the lists generated by the proposed recommendation system in terms of the Cancer topic. However, combining the results of the Wilcoxon signed-rank tests and the results of Kendall's τ, it can be concluded that the Q&A item recommendation lists generated by the proposed recommendation system are consistent with the lists ranked by the experts in terms of the Cancer topic.

4.4.5 Summary

The last research question aimed to answer whether the recommendation lists of Q&A items generated by the proposed system are relevant. To answer this question, five experts were recruited to rank the recommendation lists. The Wilcoxon signed-rank tests and Kendall's τ were used to compare the differences between the Q&A recommendation lists ranked by experts and the recommendation lists generated and ranked by the proposed system. Like RQ2, the third research question then was divided into three sub-questions, each sub-question representing one topic.

The results of the tests can be found in Tables 16-21. The intraclass correlation coefficient (ICC) was used to ensure the relevance scores data collected from five experts are reliable. The ICC scores for Q&A items Q1-Q10 from Topic Infection and Topic Heart are larger than 0.75, indicating an excellent inter-rater agreement among five experts. The ICC scores for Q11, Q14, and Q15 for Topic Cancer are larger than 0.6, suggesting a good or an excellent agreement among five experts. The ICC scores for Q12 and Q13 are larger than 0.5 but smaller than 0.75, indicating a fair agreement among experts. Therefore, the inter-rater agreement for RQ3 is either excellent, good, or fair. It concludes that the relevant scores data collected for RQ3 are reliable.

A null hypothesis of the Wilcoxon signed-rank test for Q14 failed to be rejected. It suggests that there are no significant differences between the Q&A item recommendation list generated from the proposed system and the list generated randomly. However, the mean of the relevance score of the recommendation list generated from the proposed system (2.99) is larger than the mean of the recommendation list generated randomly (1.25). It indicated that the recommendation list generated from the proposed system is more relevant than the recommendation list generated randomly, although not statistically significant. The null hypotheses of Wilcoxon signed-rank tests of other fourteen Q&A items from three topics were rejected. It suggested that the Q&A items recommendation lists generated from the proposed system are more relevant than the recommendation lists generated randomly.

The Kendall's τ coefficients for all the 15 Q&A items are larger than 0.35, indicating a strong association between the recommendation list ranked by experts and the list ranked by the proposed system.

Table 22 summarized the findings for RQ3. The null hypotheses of the correlation test for the articles Q9, Q12 and Q14 failed to be rejected. The other null hypotheses of the correlation tests for the other twelve articles were rejected.

In summary, it can be concluded the Q&A item recommendation lists generated from the proposed system are consistent with the lists ranked by the experts.

Research Questions	Hypothesis	Test	Results
RQ3.1	H_{07}	Correlation test	Q1: Rejected
			Q2: Rejected
			Q3: Rejected
			Q4: Rejected
			Q5: Rejected
RQ3.2	H_{08}	Correlation test	Q6: Rejected
			Q7: Rejected
			Q8: Rejected
			Q9: Not rejected
			Q10: Rejected
RQ3.3	H_{09}	Correlation test	Q11: Rejected
			Q12: Not rejected
			Q13: Rejected
			Q14: Not rejected
			Q15: Rejected

Table 22. Summary of findings for RQ3

4.5 Summary of results

Research questions 1, 2, and 3 were addressed by a series of statistical analyses. Table 23 summarizes the statistical findings. It suggested that the null hypotheses H_{03}, H_{04}, H_{06}, and H_{07} were rejected. The null hypotheses H_{01}, H_{02}, H_{05}, H_{08} and H_{09} failed to be rejected.

Research questions	Sub-questions	Hypothesis	Results
RQ1: Are the topic recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts?	RQ1.1 Is the topic recommendation list generated by the proposed recommendation system consistent with the list ranked by the experts in terms of the Infection topic?	H_{01}	Not rejected
	RQ1.2 Is the topic recommendation list generated by the proposed recommendation system consistent with the list ranked by the experts in terms of the Heart topic?	H_{02}	Not rejected
	RQ1.3 Is the topic recommendation list generated by the proposed recommendation system consistent with the list ranked by the experts in terms of the Cancer topic?	H_{03}	Rejected
RQ2: Are the article recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts?	RQ2.1 Are the article recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts in terms of the Infection topic?	H_{04}	Rejected
	RQ2.2 Are the article recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts in terms of the Heart topic?	H_{05}	Not rejected
	RQ2.3 Are the article recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts in terms of the Cancer topic?	H_{06}	Rejected
RQ1: Are the Q&A item recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts?	RQ3.1 Are the Q&A item recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts in terms of the Infection topic?	H_{07}	Rejected
	RQ3.2 Are the Q&A item recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts in terms of the Heart topic?	H_{08}	Not rejected
	RQ3.3 Are the Q&A item recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts in terms of the Cancer topic?	H_{09}	Not rejected

Table 23. Statistical findings for RQ 1, RQ2, and RQ3

Chapter 5. Discussion and Implications

This chapter includes a discussion of the impact of different recommendation list sizes at the three levels, the topic level, the article level, and the Q&A item level. The results of this study were compared with the cosine similarity method. The methods adopted in this study were compared with the methods applied in the previous studies. The theoretical and practical implications of the study are also discussed.

5.1 Discussion

5.1.1 The impact of the recommendation list size

In the study, the recommendation lists for topics include ten topics, whereas the recommendation lists for both articles and Q&A items include fifteen articles/Q&A items. However, different lengths of recommendation lists can produce different recommendation qualities (Peker & Koçyiğit, 2016). In this section, three sizes of topic recommendation lists, with ten topics, with eight topics, and with six topics were compared in terms of intraclass correlation coefficient (ICC) and Kendall's τ. Four sizes of article recommendation list, with fifteen articles, with thirteen articles, with eleven articles, and with 9 articles were compared. Four sizes of Q&A item recommendation lists, with fifteen items, with thirteen items, with eleven items, and with 9 items were also compared.

5.1.1.1 The impact of recommendation list at the topic level

Intraclass correlation (ICC) was used to ensure the inter-coder reliability of the five experts. Figure seventeen shows the ICC scores of three topics for three recommendation list sizes. The *Y*-axis is the value of the ICC score while the *X*-axis stands for the three topics. For

Topic Infection and Topic Cancer, the ICC scores are similar at different list sizes. For Topic Infection, the ICC score equals 0.532, 0.512, and 0.545 when the size of the recommendation lists equal ten, eight, and six respectively. For Topic Cancer, the ICC scores equal 0.449, 0.357, and 0.347 when the size of the recommendation lists equal ten, eight, and six. For Topic Heart, the highest ICC score equals 0.73 when the size of the recommendation list is 6, the lowest ICC score equaled 0.512 when the size of the list is ten. For Topic Cancer, the highest ICC score equals 0.449 when the size of the recommendation list was ten, the lowest ICC score equals 0.347 when the size of the list is 6. For Topic Infection, the highest ICC score equaled 0.545 when the size of the recommendation list was six, the lowest ICC score equaled 0.563 when the size of the list was ten. The average ICC values of the three topics were 0.51, 0.49 and 0.54 when the sizes of the list were ten, eight, and six respectively. The average ICC scores of Topic Infection was 0.530. The average ICC score of Topic Heart was 0.627 while the average ICC scores of Cancer Infection was 0.385. The ICC scores are a poor agreement (less than 0.4) when the sizes of the recommendation lists are eight and sox for Topic Cancer. As a result, a topic recommendation list with ten items is recommended in order to ensure the reliability of the study at the topic level.

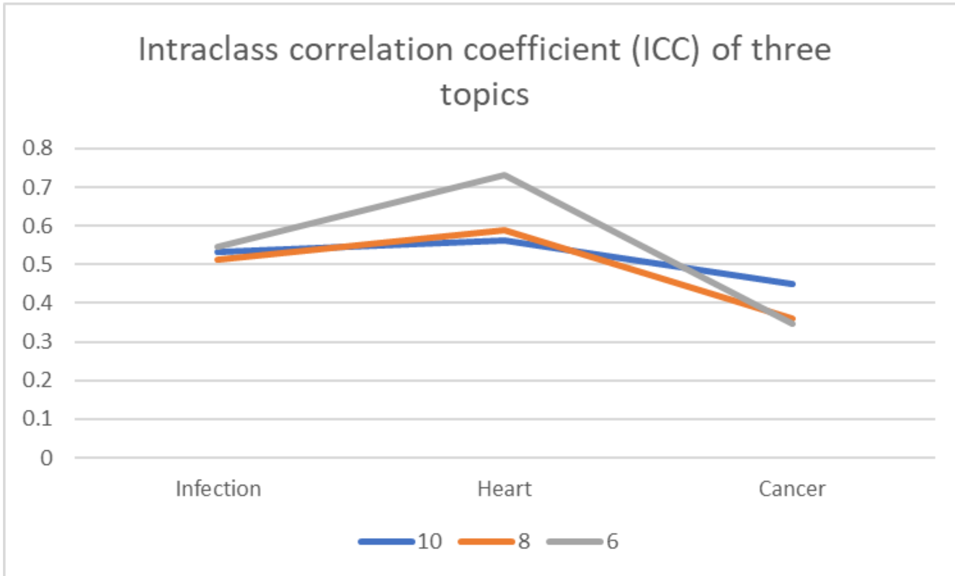


Figure 17. The intraclass correlation coefficient of the three topics for different list sizes

Kendall's τ was used to test the differences between the recommendation list generated from the proposed system and the list ranked by experts. The larger the τ is, the more similar the two lists are. In order to provide better recommendation quality, it's crucial to know which list size produced the largest Kendall's τ. Figure 18 shows Kendall's τ of three topics for three recommendation lists sizes. The *Y*-axis is the value of Kendall's τ while the *X*-axis stands for the three topics. The average Kendall's τ are 0.49, 0.50, and 0.64 when the sizes of the recommendation lists are ten, eight, and six respectively. It indicated that the shorter the recommendation list is, the better recommendation quality achieved.

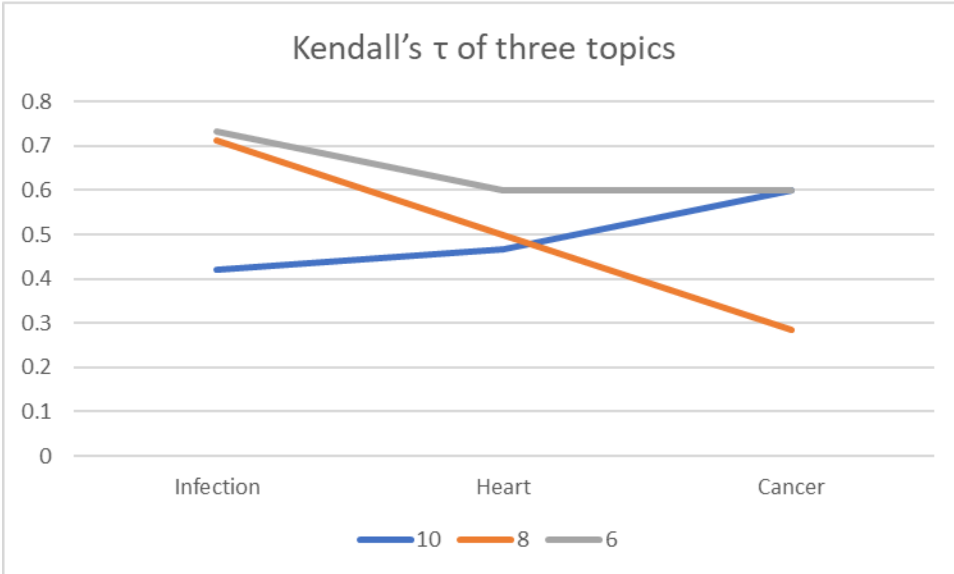


Figure 18. Kendall's τ of the three topics for different list sizes

5.1.1.2 The impact of recommendation list at the article level

Figures 19-21 display the interclass correlation (ICC) of articles from three topics. In Figures 19-21, the *Y*-axis is the value of the ICC while the *X*-axis stands for the articles. The average ICC of Topic Infection's articles was 0.718, 0.74, 0.772, and 0.786 when the recommendation list sizes are fifteen, thirteen, eleven, and nine respectively. The average ICC of Topic Heart's articles is 0.6542, 0.668, 0.693, and 0.7336 when the recommendation list sizes are fifteen, thirteen, eleven, and nine respectively. The average ICC of Topic Cancer's articles was 0.6628, 0.7144, 0.7150, and 0.458 when the recommendation list sizes were fifteen, thirteen, eleven, and nine respectively. It's not difficult to notice that the ICC scores of three topics followed the same pattern: the smaller the recommendation list, the larger the ICC value. Since the ICC value determines the agreement among the experts, it suggests that with the shorter recommendation list, the reliability of the study is better.

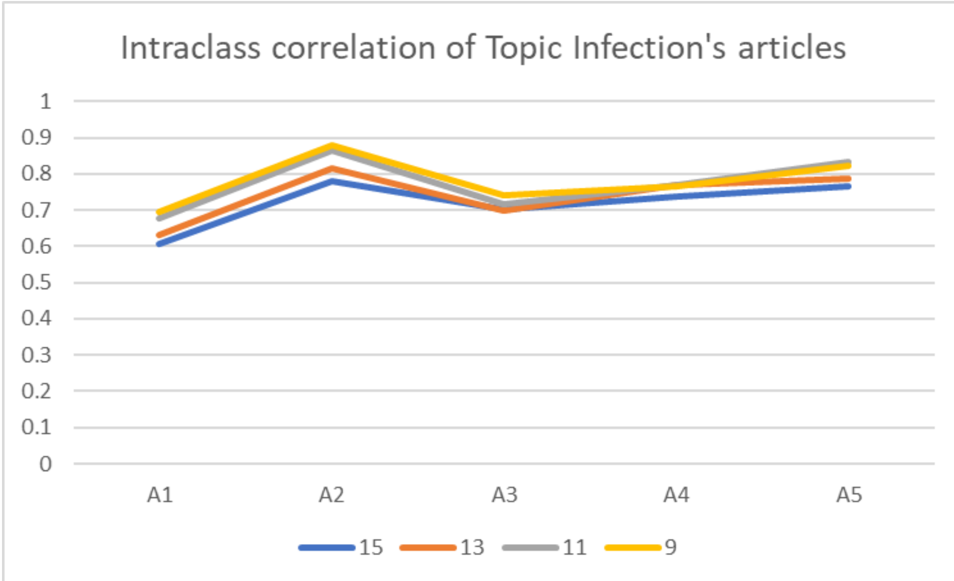


Figure 19. The intraclass correlation coefficients of Topic Infection's articles for different list sizes

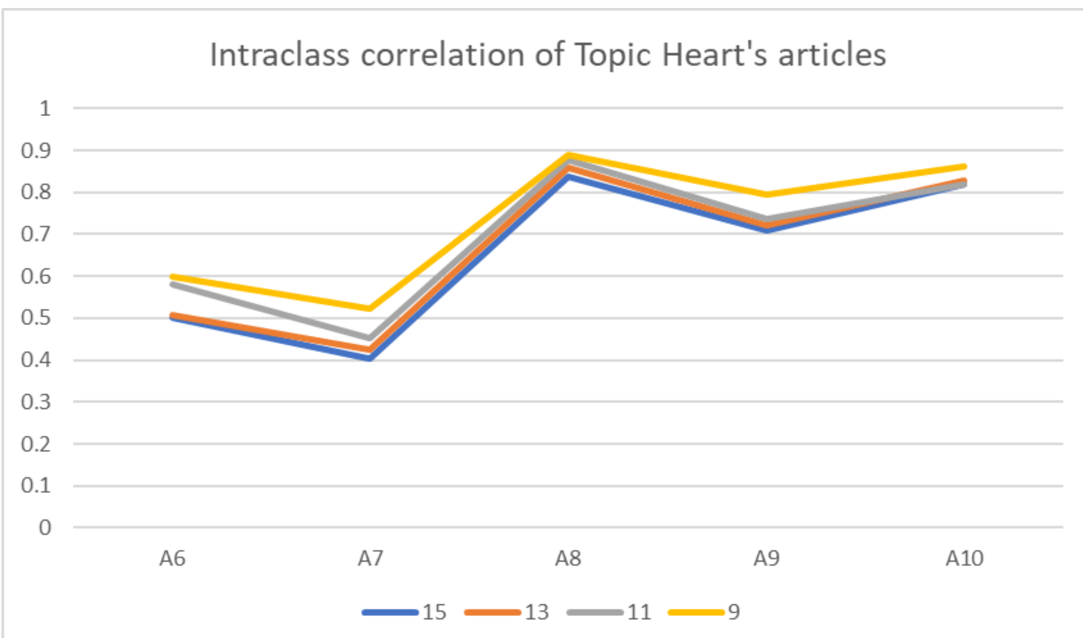


Figure 20. The intraclass correlation coefficients of Topic Heart's articles for different list sizes

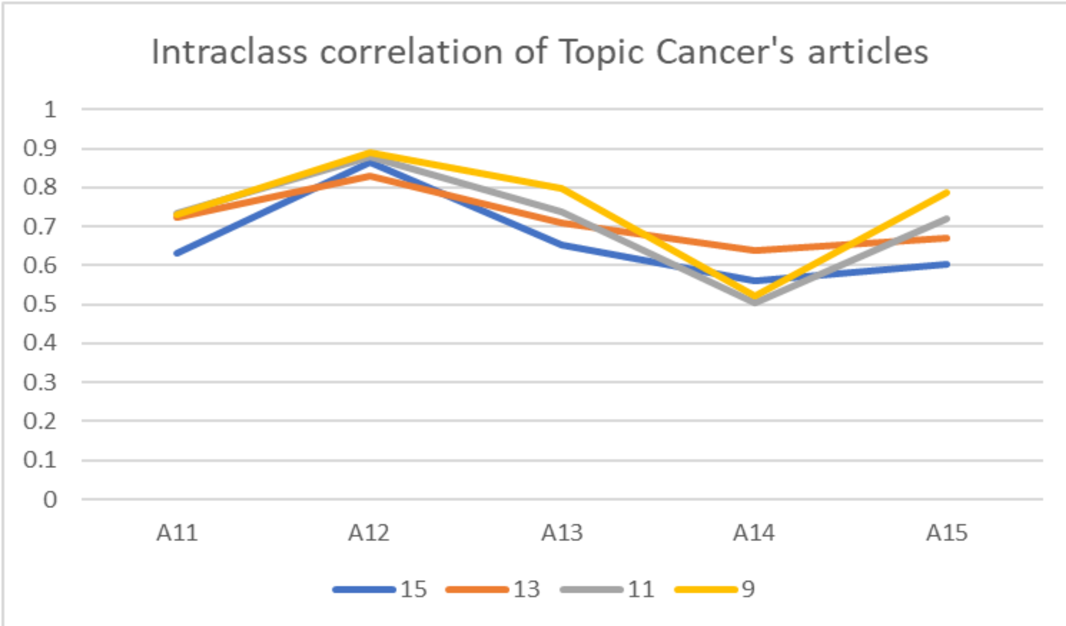


Figure 21. The intraclass correlation coefficient of Topic Cancer's articles for different list sizes

Figures 22-24 show Kendall's τ of articles from the three topics. In Figures 22-24, the *Y*-axis is the value of Kendall's τ while the *X*-axis stands for the articles. The average Kendall's τ of Topic Infection was 0.5962, 0.5282, 0.491, and 0.589 when the recommendation list size was fifteen, thirteen, eleven, and nine respectively. It shows the largest average Kendall's τ is 0.5962 with the longest recommendation list, while the smallest average Kendall's is 0.491 when the recommendation list size is eleven. The average Kendall's τ of Topic Heart is 0.5316, 0.5230, 0.5128, and 0.7222 when the recommendation list size is fifteen, thirteen, eleven, and nine respectively. It clearly shows Kendall's τ performed best with the shortest recommendation list. However, for Topic Cancer, Kendall's τ performed worst with the shortest recommendation list. The average Kendall's τ of Topic Cancer is 0.602, 0. 4462, 0.5054, and 0.3668 when the recommendation list size is fifteen, thirteen, eleven, and nine respectively.

As a result, there are no consistent patterns of the impact of the size of the recommendation list at the article level on Kendall's τ.

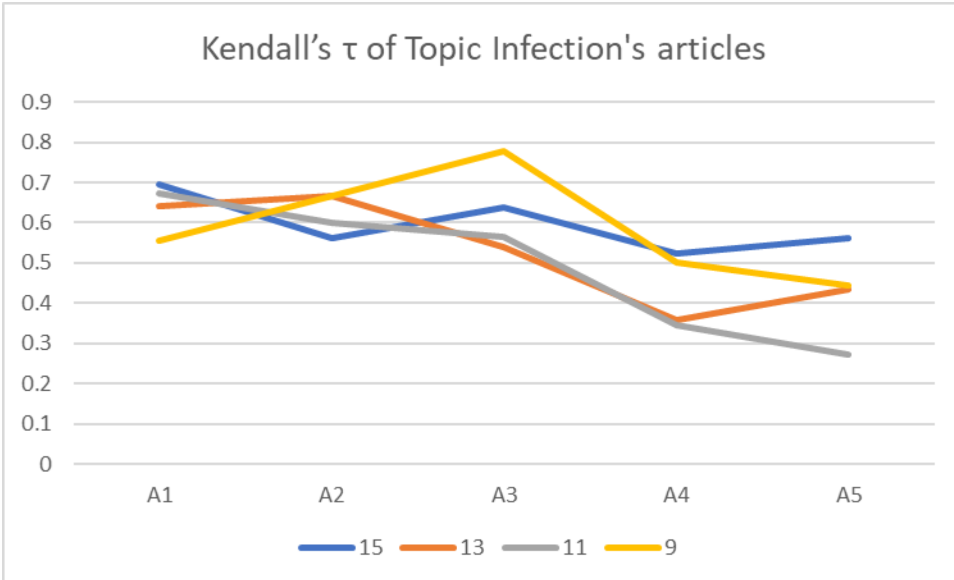


Figure 22. Kendall's τ of Topic Infection's articles for different list sizes

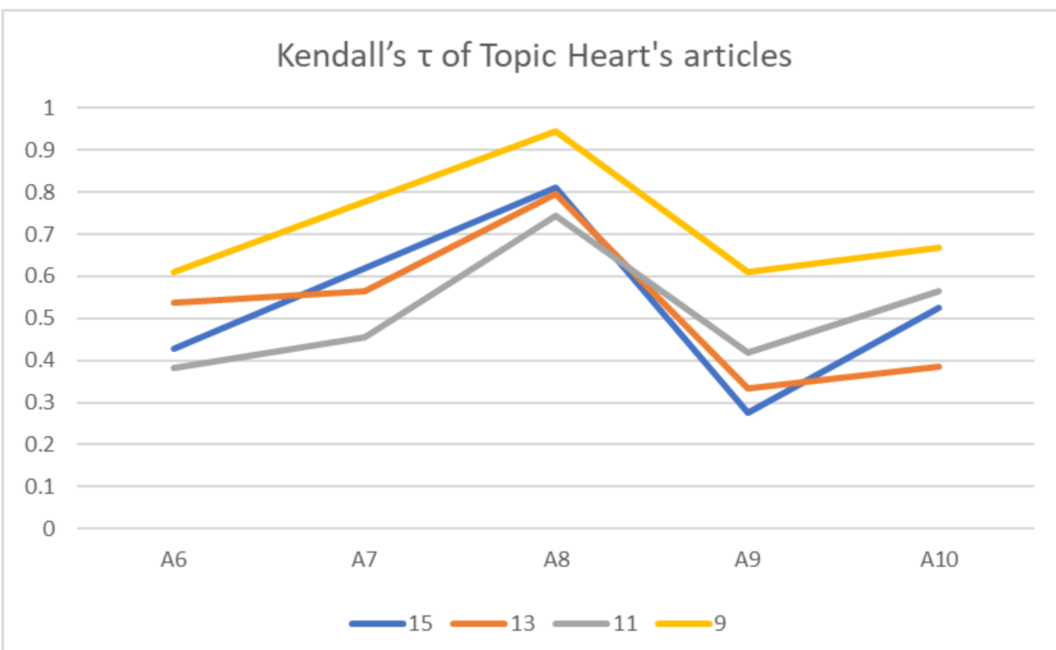


Figure 23. Kendall's τ of Topic Infectionss articles for different list sizes

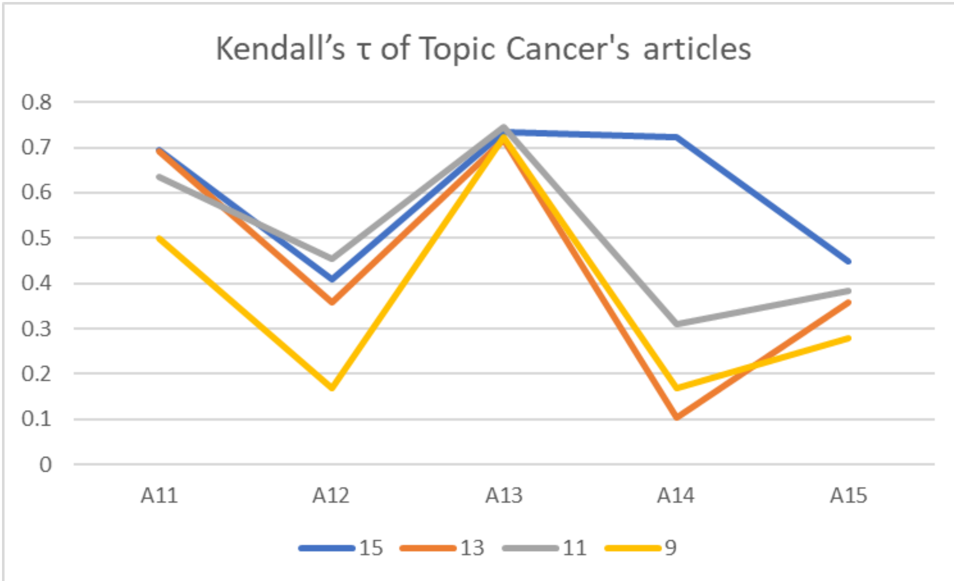


Figure 24. Kendall's τ of Topic Infection's articles for different list sizes

5.1.1.3 The impact of recommendation list at Q&A item level

Figures 25-27 shows the interclass correlation (ICC) of Q&A items from three topics. In Figures 25-27, the *Y*-axis is the value of ICC while the *X*-axis stands for the Q&A items. The average ICC scores of Topic Infection's Q&A items were 0.756, 0.752, 0.772, 0.84 when the recommendation list sizes were fifteen, thirteen, eleven, and nine respectively. The average ICC scores of Topic Heart's Q&A items were 0.8602, 0.8718, 0.8812, and 0.89 when the recommendation list sizes were fifteen, thirteen, eleven, and nine respectively. The average ICC scores of Topic Cancer's Q&A items were 0.6814, 0.6902, 0.7212, and 0.7508 when the recommendation list sizes are fifteen, thirteen, eleven, and nine respectively. It is easy enough to notice that the ICC scores of the three topics followed the same pattern: the smaller the recommendation list, the larger the ICC value. Since the ICC value determines the agreement among the experts, it suggests that a shorter recommendation list results in better reliability of the study.

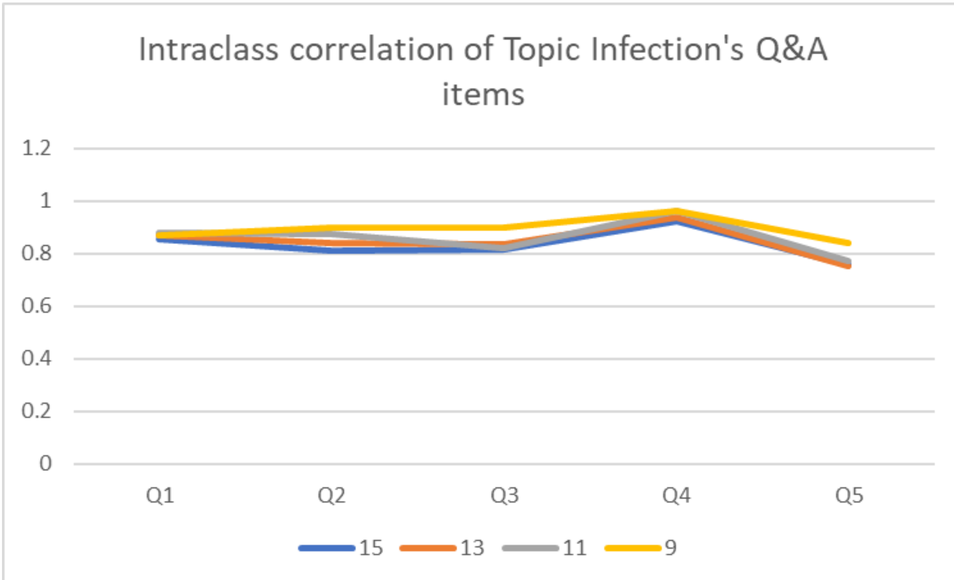


Figure 25. The intraclass correlation coefficient of Topic Infection's Q&A items for different list sizes

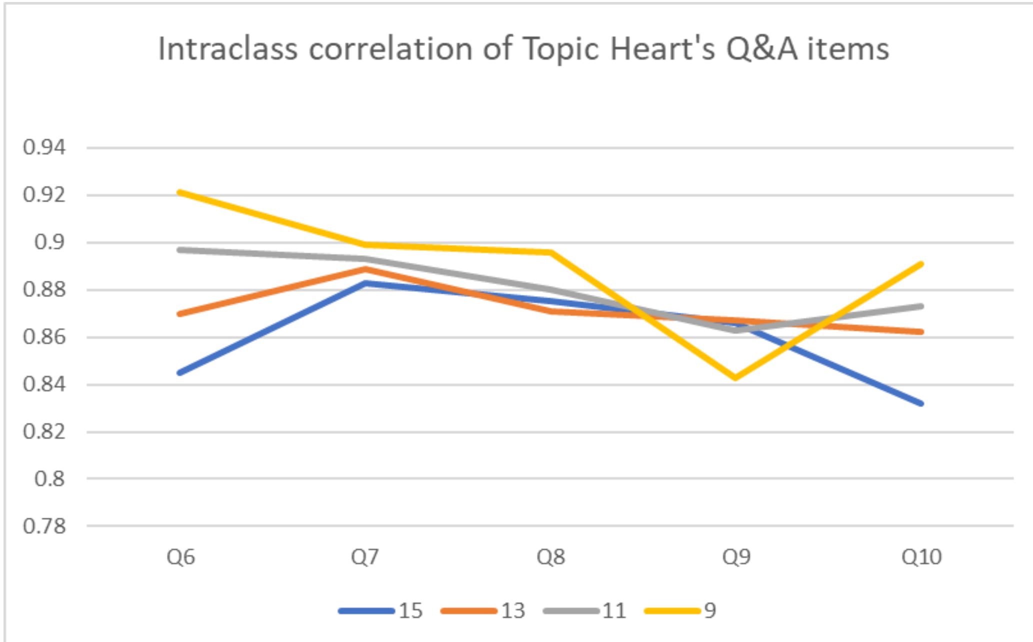


Figure 26. The intraclass correlation coefficient of Topic Heart's Q&A items for different list sizes

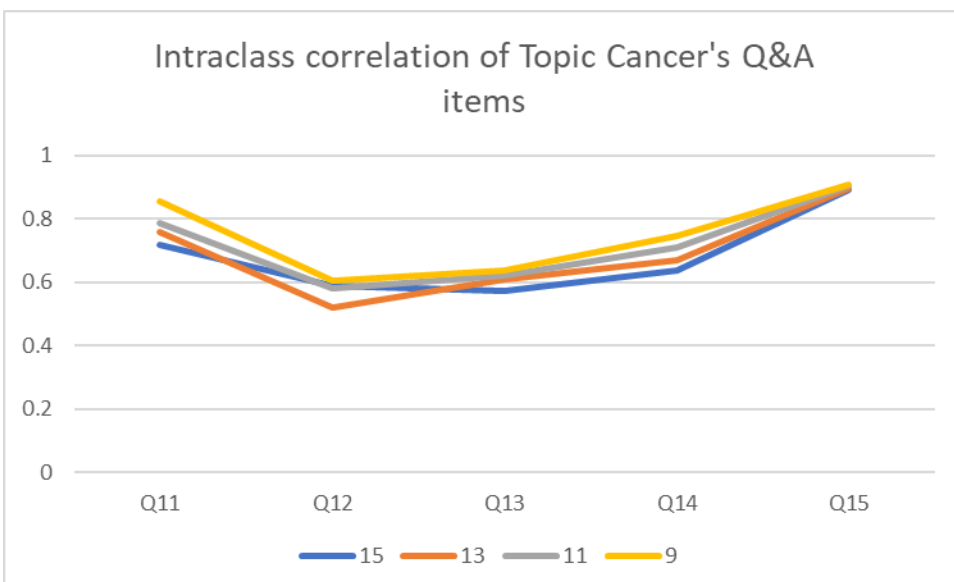


Figure 27. The intraclass correlation coefficient of Topic Cancer's Q&A items for different list sizes

Figures 28-30 show Kendall's τ of Q&A items from three topics. In Figures 28-30, the *Y*-axis is the value of Kendall's τ while the *X*-axis stands for the Q&A items. The average Kendall's τ of Topic Infection is 0.6458, 0.6206, 0.5926, and 0.6444 when the recommendation list size is fifteen, thirteen, eleven, and nine respectively. The average Kendall's τ of Topic Heart is 0.3522, 0.314, 0.33, and 0.278 when the recommendation list size is fifteen, thirteen, eleven, and nine respectively. The average Kendall's τ of Topic Cancer is 0.5388, 0.5692, 0.5636, and 0.5998 when the recommendation list size is fifteen, thirteen, eleven, and nine respectively. Like the results of Kendall's τ at the article level, there are no consistent patterns of the impact of the size of the recommendation list at the Q&A item level on Kendall's τ.

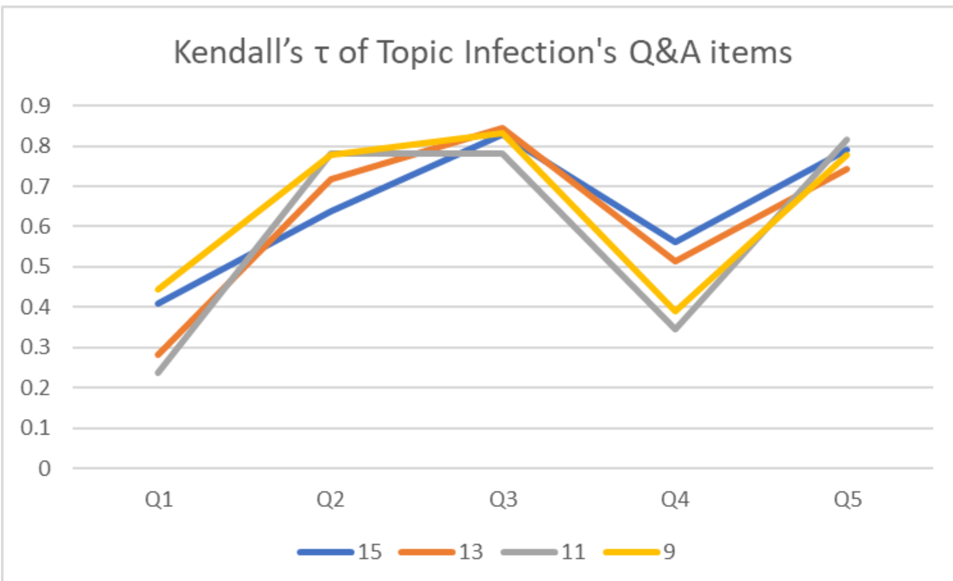


Figure 28. Kendall's τ of Topic Infection's Q&A items for different list sizes

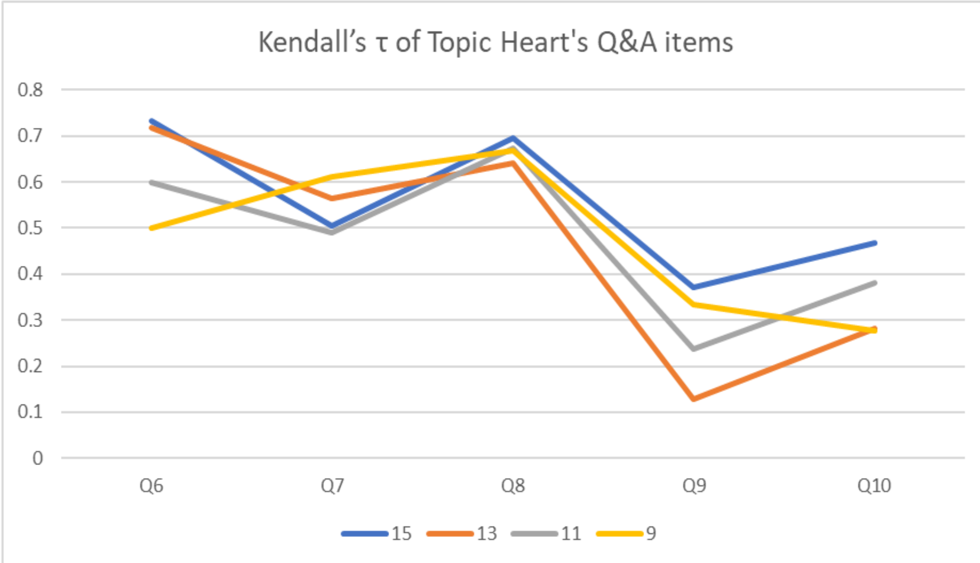


Figure 29. Kendall's τ of Topic Heart's Q&A items for different list sizes

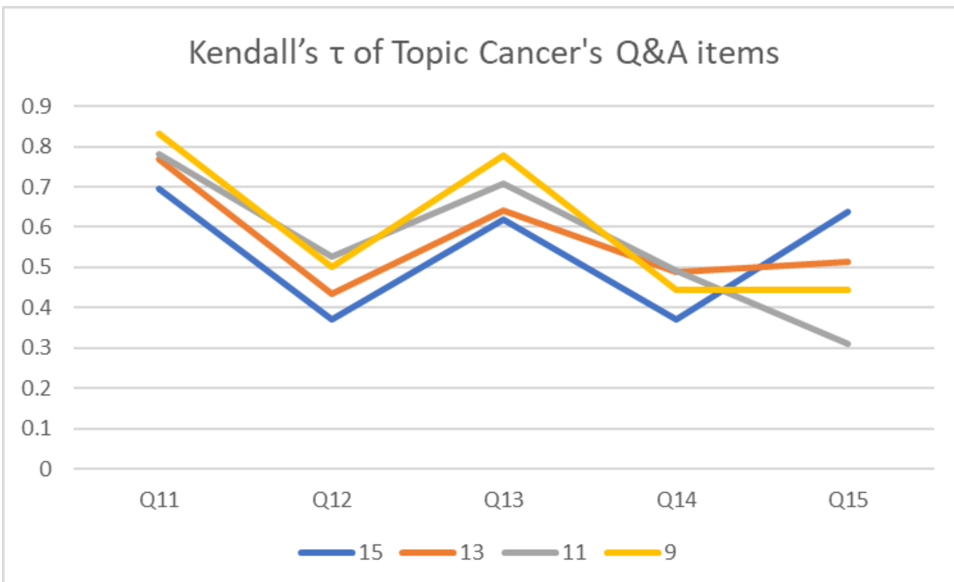


Figure 30. Kendall's τ of Topic Cancer's Q&A items for different list sizes

In a summary, although the recommendation quality is the best with the shortest recommendation list at the topic level in terms of Kendall's τ, there are no consistent patterns of the impact of the size of the recommendation list at the article level and the Q&A item level in terms of Kendall's τ in this study. In the study, only articles and Q&A items from Topic Cancer were analyzed for the impact of the size of recommendation lists. If more articles and Q&A items had been included, patterns would have been revealed. In a future study, more articles, and Q&A items from other topics could be analyzed to uncover patterns.

5.1.2 The comparison with the cosine similarity method

This study compared the recommendation list generated from the proposed system and the recommendation list ranked by the experts. It showed that the proposed system can generate a comparable recommendation list. To better explore the performance of the proposed recommendation system from a different angle, the recommendation lists ranked by the cosine similarity method were generated and compared with the recommendation lists ranked by the experts. Kendall's τ results from the new similarity method are obtained to evaluate the performance of the cosine similarity method. The topic Heart was selected for the comparison between the cosine method and the proposed method in this study. At the Topic level, Kendall's τ obtained from the proposed system is 0.467 whereas obtained from the cosine similarity method is 0.511. The cosine similarity method performed better.

Figure 31 shows the comparison of the article level between the two methods in terms of Kendall's τ. In Figure 31, The *Y*-axis is the value of Kendall's τ while the *X*-axis stands for the articles. The average Kendall's τ of the proposed system is 0.5316, which is larger than the average of Kendall's τ (0.5048) of the cosine method. It suggests that the proposed method performed slightly better than the cosine similarity method.

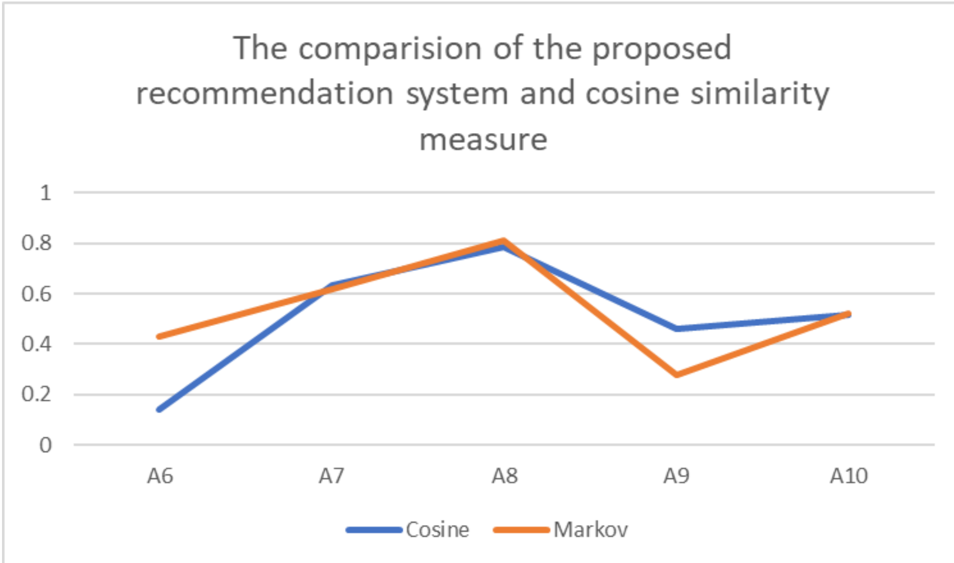


Figure 31. The comparison of the two methods at the article level.

At the Q&A items level, Figure 32 shows the comparison between the two methods. In Figure 32, The *Y*-axis is the value of Kendall's τ while the *X*-axis stands for the Q&A items. The average Kendall's τ of the proposed system was 0.5084, which is larger than the average of Kendall's τ (0.4922) from the cosine method. It indicates that the proposed method performed slightly better than the cosine similarity method.

In a summary, the cosine similarity method performed slightly better at the topic level. The proposed method performed slightly better at both the article level and Q&A level. It is concluded that the proposed method performed similarly to the cosine similarity method. Although these two methods performed similarly. the cosine similarity methods provide a fixed recommendation list, whereas the Markov chain offers a dynamic recommendation list that can be updated based on the recent user activities or for providing a customized or personal recommendation service.

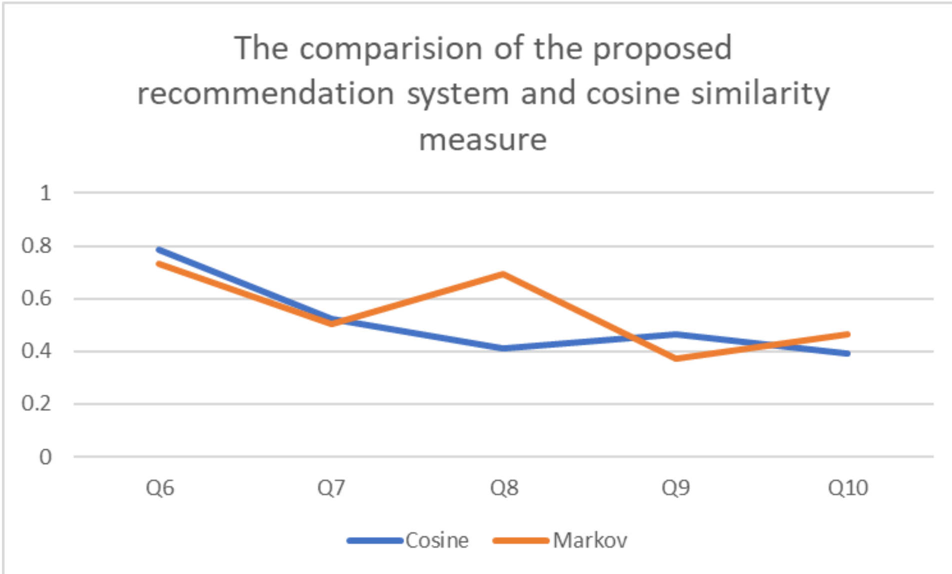


Figure 32. The comparison of the two methods at the Q&A items level.

5.1.3 The comparisons between this study and other similar studies

This study brought a unique methodology to build a recommendation system for a health portal, while previous studies applied other approaches to building recommendation systems. Traditional recommendation systems can be divided into two groups: the content-based recommendation systems (Mooney & Roy, 2000) and the collaborative filtering recommendation systems (Lai, 2003; Linden, Smith, & York, 2003; Zhou et al., 2008). The content-based systems recommend items based on the features of the items, whereas the collaborative filtering systems recommend items based on the characteristics of the users. The proposed recommendation system in this study incorporated the Markov chain method and transaction log analysis for generating the recommendation lists. Polyzou et al. (2019) incorporated the Markov chain method for course recommendation. Sejal et al. (2016) adopted the Markov chain method for image recommendation. This study applied the Markov chain method to generate the recommendation lists at the three levels: the topic level, the article level, and the Q&A item level. Previous studies provided many ways to improve the subject directories. Zhang et al. (2009)

adopted a Self-Organizing Map (SOM) with a U-matrix algorithm to improve the visual health subject directory. Zhu and Zhang (2020) applied the social network analysis to improve the MedlinePlus subject directory. This study proposed a recommendation system to improve the usability of the subject directory in addition to the traditional query search and browsing mechanism.

5.2 Implications

5.2.1 Theoretical implications

Previous studies on the information patterns of health portal users adopted qualitative research methods including interviews, "think aloud" protocols, and focused groups to gather qualitative data. This study extracted traversal data from a transaction log that recorded users' traversal activities on the portal. The traversal data extracted from the transaction log can help researchers better understand portal users in addition to the qualitative data.

One of the most recognized issues of the Markov chain analysis method is how to deal with sparse matrices. In this study, a method of combining the traversal data matrices with the Cosine-similarity based matrices was proposed. A coefficient of 0.15 was set to adjust the influence of the Cosine-similarity matrix while the coefficient of the traversal data matrix was 0.85.

Previous studies applied numerous methods to improve Web-based subject directories. This study proposed a recommendation system for a health portal subject directory by adopting the Markov chain method. The traditional recommendation systems focused on recommending general items to users, which include articles, images, journal papers, commodities, movies, etc. The proposed recommendation system can recommend information at three levels, the topic level, the article level, and the Q&A item level. The proposed recommendation system can also facilitate users to find relevant information in addition to the traditional query search method and the browsing mechanism.

5.2.2 Practical implications

This study proposed a recommendation system for the HealthLink portal, which didn't particularly have a recommendation mechanism. The recommendation system facilitated health consumers to find better-related information while browsing Web pages. The proposed system offered recommended lists of topics, articles, and Q&A items. Users can obtain relevant information without searching for anything.

The proposed recommendation procedures and methods can also be integrated into other types of portals, including government or official portals, agriculture-related portals, sport-related portals, entertainment portals, and so on. In fact, any information portal with a subject directory as well as full-text articles can take advantage of the proposed recommendation procedures and methods.

It improved the effectiveness of user navigation on a portal. Users can not only follow the existing navigation routes offered by the system, but also click the recommended topics, articles, and Q&A items to browse for more relevant information.

Another significant application of the study findings is to enhance users' search experience in addition to query search. An internal search engine is prevalent in many public portals for helping users locate needed information. The proposed recommendation procedures and methods offered more related articles or items for a retrieved article on a returned retrieval list.

5.2.2 Methodological implications

Multiple methods were synthesized to analyze how to effectively recommend related and relevant information on a health portal. This study used four methods: the transaction logs analysis method, the Markov chain analysis method, the relevance judgment method, and the inferential analysis method. The transaction log analysis was adopted to understand the users' traversal activities from the users' perspective on the portal. The Markov chain analysis method was applied to generate the recommendation lists based on the users' browsing data. The relevance judgment method in addition to the inferential analysis method was utilized to determine the consistency of the recommendation lists generated from the proposed system and the lists ranked by the experts. The proposed mixed methodology can be applied to examine issues of other online information portals, such as government or official portals, agriculture-related portals, sport-related portals, entertainment portals, and so on.

5.3 Summary

The results of this study were compared with the results from different recommendation list sizes. The size of topic recommendation list was ten, the size of both article recommendation list and Q&A item recommendation was fifteen in this study. It was concluded that although the recommendation quality is the best with the shortest recommendation list at the topic level, there are no consistent patterns of the impact of the size of the recommendation list at the article level and the Q&A item level. The results of this study were compared with the results from the cosine similarity method, it was concluded that the proposed method performed similarly to the cosine similarity method. By comparing with the previous studies, this study brought a unique way to build a recommendation system for a health portal.

The theoretical implications and practical implications were discussed. The proposed mixed method in this study can be applied to investigate other online information portals by constructing a recommendation system.

Chapter 6. Conclusions

The final chapter summarizes the research questions and the findings of the study. The limitation and future directions are also discussed.

6.1 Summary of research questions

The primary research problem of this study was to investigate whether the recommended information generated by the Markov Chain method based on users' traversal activities on the HealthLink portal is effective. Toward this aim, a mixed-method which included the transaction log analysis method, the Markov chain analysis method, and the inferential analysis method was employed to analyze the collected data.

RQ1: Are the topic recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts?

The first research question explored the traversal activities between nodes in the subject directory. The significance of RQ1 was to investigate whether the recommended topics generated by the Markov Chain method based on users' traversal activities on the HealthLink portal are effective. RQ1.1 investigated the consistency between the topic recommendation list generated by the proposed method and the list ranked by experts in terms of the Infection topic, while RQ1.2 examined the Heart Infection topic recommendation list and RQ1.3 explored the Cancer topic recommendation list. The null hypothesis of RQ1.3 was rejected. It suggested that the recommendation list generated from the proposed system is highly consistent with the list ranked by experts in terms of the Cancer topic. Although the null hypotheses of RQ1.1 and RQ1.2 failed to be rejected. Strong associations between the lists generated by the proposed system and the lists ranked by experts were found through the values of Kendall's τ. It indicated that the recommendation lists generated from the proposed system are consistent with the lists ranked by experts in terms of the Infection topic and the Heart topic. As a result, it was concluded that the topic recommendation lists generated by the proposed recommendation system are consistent with the lists ranked by the experts.

RQ2: Are the article recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts?

RQ2 investigated the traversal activities between articles in the subject directory. The significance of RQ2 was to investigate whether the recommended articles generated by the Markov Chain method based on users' traversal activities on the HealthLink portal are effective. RQ2.1 explored the consistency between the article recommendation lists generated by the proposed method and the lists ranked by experts in terms of the Infection topic, while RQ2.2 examined the Heart Infection article recommendation lists and RQ2.3 examined the Cancer article recommendation lists. The null hypotheses of RQ2.1 and RQ2.3 were rejected. It indicated that the article recommendation lists generated from the proposed system are highly consistent with the lists ranked by experts in terms of the Infection topic and the Cancer topic. The null hypothesis of RQ2.2 failed to be rejected. Associations between the article recommendation lists generated by the proposed system and the lists ranked by experts were discovered through the values of Kendall's τ. Therefore, it was concluded that the article recommendation lists generated by the proposed recommendation system are consistent with the lists ranked by the experts.

RQ3: Are the Q&A item recommendation lists generated by the proposed recommendation system consistent with the lists ranked by the experts?

The last research question examined the traversal activities between Q&A items in the subject directory. RQ3.1 examined the consistency between the Q&A item recommendation lists generated by the proposed method and the lists ranked by experts in terms of the Infection topic, while RQ3.2 explored the Heart Infection Q&A item recommendation lists and RQ3.3 examined the Cancer Q&A recommendation lists. The null hypothesis of RQ3.1 was rejected, which indicated that the Q&A item recommendation lists generated from the proposed system are highly consistent with the lists ranked by experts in terms of the Infection topic. The null hypotheses of RQ3.2 and RQ3.3 failed to be rejected. However, strong or medium associations between the Q&A item recommendation lists generated by the proposed system and the lists ranked by experts were found through the values of Kendall's τ. As a result, it was concluded that the Q&A item recommendation lists generated by the proposed recommendation system are consistent with the lists ranked by the experts.

6.2 Limitations

There are certain limitations to the study. The limitations include, but are not limited to: the use of only the HealthLink portal, the limited timeframe of the transaction log used, the number of topics, articles, and Q&A items used to test the proposed recommendation system, the lack of user study, and the only use of a first-order Markov chain.

The first limitation to this study related to the sampling and data collection. The HealthLink portal might not be representative of all health portals. The transaction log data collected was set as 1 year. The time selected might not be able to present a whole picture of users’ s traversal pattern. The number of topics used was three out of forty-seven total topics on the portal, which were Topic Infection, Topic Heart, and Topic Cancer. The number of articles and Q&A items used was both fifteen, which may not present the entire data on the portal.

Another limitation to this study is that the recommendation system was constructed by a first-order Markov chain that only considered the current state. Higher-order Markov chains which also consider the previous states were not examined in the study. The higher order Markov chain analysis requires higher computational power. For instance, the topic transition matrix generated for this study is a 47*47 matrix. If the second order Markov chain is adopted, the same topic Markov chain matrix would have 47*47*47 cells, which would bring much more computational burden to the study. The higher order Markov chain offers more detailed user traversal information, it will be considered for future research.

The last limitation to this study is related to the research methods used. This study adopted only quantitative research methods to analyze the data collected. Adopting certain qualitative methods could provide a clearer picture of users’ perspectives. Qualitative methods such as interviews or “think aloud” protocols were not considered in this study. However, to better build a recommendation system, qualitative methods should be conducted in the future.

6.3 Future directions

As discussed above, this study centered on the quantitative data collected from the transaction log without considering a user study. A future study that includes interviewing the health portal users as well as recruiting them to use the recommendation system may bring more findings.

In addition to the methodology, a larger scale of quantitative data should be collected to validate the results of this study. More portals such as government or official portals, agriculture-related portals, sport-related portals, and entertainment portals should be considered. The proposed mixed method in this study can be applied to these portals to test the performance of the recommendation system in future studies. More topics, articles, and Q&A items that can better represent the portal should be considered. The transaction log data that contains a long time should be considered. Results from the transaction log that contain a long time might further discover users' traversal patterns.

This study compared the proposed method with the cosine similarity method, a future study that includes more similarity methods such as topic modeling may bring more interesting findings. This study adopted a first-order Markov chain that only considers the current state, a future study that adopted higher-order Markov chains that also consider the previous states may bring richer findings.

References

Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, 207–216. https://doi.org/10.1145/170035.170072

Ajiferuke, I., & Wolfram, D. (2004). Informetric modelling of internet search and browsing characteristics. *Informetric Modelling of Internet Search and Browsing Characteristics*, *28*(1), 1–16.

Ajiferuke, I., Wolfram, D., & Famoye, F. (2006). Sample size and informetric model goodness-of-fit outcomes: A search engine log case study. *Journal of Information Science*, *32*(3), 212–222. https://doi.org/10.1177/0165551506064361

Ajiferuke, I., Wolfram, D., & Xie, H. (Iris). (2004). Modelling Website Visitation and Resource Usage Characteristics by IP Address Data. *Proceedings of the Annual Conference of CAIS / Actes Du Congrès Annuel de l'ACSI*. https://doi.org/10.29173/cais315

Beeferman, D., & Berger, A. (2000). Agglomerative clustering of a search engine query log. *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 407–416. https://doi.org/10.1145/347090.347176

Beitzel, S. M., Jensen, E. C., Chowdhury, A., Frieder, O., & Grossman, D. (2007). Temporal analysis of a very large topically categorized Web query log. *Journal of the American Society for Information Science and Technology*, *58*(2), 166–178. https://doi.org/10.1002/asi.20464

Benevenuto, F., Rodrigues, T., Cha, M., & Almeida, V. (2009). Characterizing user behavior in online social networks. *Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement*, 49–62.

Benoît, G. (2005). Application of Markov chains in an interactive information retrieval system. *Information Processing & Management*, *41*(4), 843–857.

Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer.

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network Analysis in the Social Sciences. *Science*, *323*(5916), 892–895. https://doi.org/10.1126/science.1165821

Borges, J., & Levene, M. (2004). A dynamic clustering-based Markov model for web usage mining. *ArXiv Preprint Cs/0406032*.

Borges, J., & Levene, M. (2005). Generating Dynamic Higher-Order Markov Models in Web Usage Mining. In A. M. Jorge, L. Torgo, P. Brazdil, R. Camacho, & J. Gama (Eds.), *Knowledge Discovery in Databases: PKDD 2005* (pp. 34–45). Springer. https://doi.org/10.1007/11564126_9

Boyd, D. M., & Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, *13*(1), 210–230. https://doi.org/10.1111/j.1083-6101.2007.00393.x

Brin, S., & Page, L. (2012). Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer Networks*, *56*(18), 3825–3833.

Cadez, I., Heckerman, D., Meek, C., Smyth, P., & White, S. (2000). Visualization of navigation patterns on a web site using model-based clustering. *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 280–284.

Chakrabarti, S., Dom, B., & Indyk, P. (1998). Enhanced hypertext categorization using hyperlinks. *ACM SIGMOD Record*, *27*(2), 307–318. https://doi.org/10.1145/276305.276332

Chen, M. Y., Kundu, A., & Zhou, J. (1994). Off-line handwritten word recognition using a hidden Markov model type stochastic network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *16*(5), 481–496. https://doi.org/10.1109/34.291449

Chen, S. Y., Magoulas, G. D., & Dimakopoulos, D. (2005). A flexible interface design for Web directories to accommodate different cognitive styles. *Journal of the American Society for Information Science and Technology*, *56*(1), 70–83. https://doi.org/10.1002/asi.20103

Chung, W., Bonillas, A., Lai, G., Xi, W., & Chen, H. (2006). Supporting non-English Web searching: An experiment on the Spanish business and the Arabic medical intelligence portals. *Decision Support Systems*, *42*(3), 1697–1714. https://doi.org/10.1016/j.dss.2006.02.015

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment, 6(4), 284.

Cline, R. J. W., & Haynes, K. M. (2001). Consumer health information seeking on the Internet: The state of the art. *Health Education Research*, *16*(6), 671–692. https://doi.org/10.1093/her/16.6.671

Coady, R. P. (1983). Testing for Markov-chain properties in the circulation of science monographs. *Information Processing & Management*, *19*(5), 279–284.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement, 20(1), 37–46. https://doi.org/10.1177/001316446002000104

Cooper, A. (2008). A survey of query log privacy-enhancing techniques from a policy perspective. *ACM Transactions on the Web*, *2*(4), 19:1–19:27. https://doi.org/10.1145/1409220.1409222

Cormack, G. V., & Horspool, R. N. S. (1987). Data Compression Using Dynamic Markov Modelling. *The Computer Journal*, *30*(6), 541–550. https://doi.org/10.1093/comjnl/30.6.541

Cowie, J., & Lehnert, W. (1996). Information extraction. *Communications of the ACM*, *39*(1), 80–91.

Daniłowicz, C., & Baliński, J. (2001). Document ranking based upon Markov chains. *Information Processing & Management*, *37*(4), 623–637.

Davis, P. M. (2004). Information-seeking behavior of chemists: A transaction log analysis of referral URLs. *Journal of the American Society for Information Science and Technology*, *55*(4), 326–332. https://doi.org/10.1002/asi.10384

Deshpande, M., & Karypis, G. (2004a). Item-based top-*N* recommendation algorithms. *ACM Transactions on Information Systems*, *22*(1), 143–177. https://doi.org/10.1145/963770.963776

Deshpande, M., & Karypis, G. (2004b). Selective Markov models for predicting Web page accesses. *ACM Transactions on Internet Technology*, *4*(2), 163–184. https://doi.org/10.1145/990301.990304

Eirinaki, M., Vazirgiannis, M., & Kapogiannis, D. (2005). Web path recommendations based on page ranking and markov models. *Proceedings of the 7th Annual ACM International Workshop on Web Information and Data Management*, 2–9.

Eppler, M. J., & Mengis, J. (2008). The concept of information overload-a review of literature from organization science, accounting, marketing, mis, and related disciplines (2004). In *Kommunikationsmanagement im Wandel* (pp. 271–305). Springer.

Eyrich, N., Padman, M. L., & Sweetser, K. D. (2008). PR practitioners' use of social media tools and communication technology. *Public Relations Review*, *34*(4), 412–414. https://doi.org/10.1016/j.pubrev.2008.09.010

Fisher, N. I., Lewis, T., & Embleton, B. J. J. (1993). *Statistical Analysis of Spherical Data*. Cambridge University Press.

Flynn, K. E., Smith, M. A., & Freese, J. (2006). When Do Older Adults Turn to the Internet for Health Information? Findings from the Wisconsin Longitudinal Study. *Journal of General Internal Medicine*, *21*(12), 1295–1301. https://doi.org/10.1111/j.1525-1497.2006.00622.x

Fonseca, B.M., Golgher, P. B., de Moura, E. S., & Ziviani, N. (2003). Using association rules to discover search engines related queries. *Proceedings of the IEEE/LEOS 3rd International Conference on Numerical Simulation of Semiconductor Optoelectronic Devices (IEEE Cat. No.03EX726)*, 66–71. https://doi.org/10.1109/LAWEB.2003.1250284

Fonseca, Bruno M., Golgher, P., Pôssas, B., Ribeiro-Neto, B., & Ziviani, N. (2005). Concept-based interactive query expansion. *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, 696–703. https://doi.org/10.1145/1099554.1099726

Fox, S. (2006). *Online Health Search 2006*. Retrieved from http://www.pewinternet.org/2006/10/29/online-health-search-2006/

Frasconi, P., Soda, G., & Vullo, A. (2002). Hidden Markov models for text categorization in multi-page documents. *Journal of Intelligent Information Systems*, *18*(2–3), 195–217.

Freitag, D., & McCallum, A. (1999). Information extraction with HMMs and shrinkage. *Proceedings of the AAAI-99 Workshop on Machine Learning for Information Extraction*, 31–36.

Gazan, R. (2011). Social Q&A. *Journal of the American Society for Information Science and Technology*, *62*(12), 2301–2312. https://doi.org/10.1002/asi.21562

Göker, A., & He, D. (2000). Analysing Web Search Logs to Determine Session Boundaries for User-Oriented Learning. In P. Brusilovsky, O. Stock, & C. Strapparava (Eds.), *Adaptive Hypermedia and Adaptive Web-Based Systems* (pp. 319–322). Springer. https://doi.org/10.1007/3-540-44595-1_38

Gray, N. J., Klein, J. D., Noyce, P. R., Sesselberg, T. S., & Cantrill, J. A. (2005). Health information-seeking behaviour in adolescence: The place of the internet. *Social Science & Medicine*, *60*(7), 1467–1478. https://doi.org/10.1016/j.socscimed.2004.08.010

Han, H., & Wolfram, D. (2016). An exploration of search session patterns in an image-based digital library. *Journal of Information Science*, *42*(4), 477–491. https://doi.org/10.1177/0165551515598952

Han, J., Cheng, H., Xin, D., & Yan, X. (2007). Frequent pattern mining: Current status and future directions. *Data Mining and Knowledge Discovery*, *15*(1), 55–86. https://doi.org/10.1007/s10618-006-0059-1

Harper, F. M., Raban, D., Rafaeli, S., & Konstan, J. A. (2008). Predictors of answer quality in online Q&A sites. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 865–874. https://doi.org/10.1145/1357054.1357191

Hayes, A. F., & Krippendorff, K. (2007). Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*, *1*(1), 77–89.

He, D., Göker, A., & Harper, D. J. (2002a). Combining evidence for automatic Web session identification. *Information Processing & Management*, *38*(5), 727–742. https://doi.org/10.1016/S0306-4573(01)00060-7

He, D., Göker, A., & Harper, D. J. (2002b). Combining evidence for automatic Web session identification. *Information Processing & Management*, *38*(5), 727–742. https://doi.org/10.1016/S0306-4573(01)00060-7

Jansen, B. J. (2006). Search log analysis: What it is, what's been done, how to do it. *Library & Information Science Research*, *28*(3), 407–432. https://doi.org/10.1016/j.lisr.2006.06.005

Jansen, B. J., Booth, D. L., & Spink, A. (2008). Determining the informational, navigational, and transactional intent of Web queries. *Information Processing & Management*, *44*(3), 1251–1266. https://doi.org/10.1016/j.ipm.2007.07.015

Jansen, B. J., & McNeese, M. D. (2005). Evaluating the effectiveness of and patterns of interactions with automated searching assistance. *Journal of the American Society for Information Science and Technology*, *56*(14), 1480–1503. https://doi.org/10.1002/asi.20242

Jansen, B. J., & Pooch, U. (2001). A review of Web searching studies and a framework for future research. *Journal of the American Society for Information Science and Technology*, *52*(3), 235–246. https://doi.org/10.1002/1097-4571(2000)9999:9999<::AID-ASI1607>3.0.CO;2-F

Jansen, B. J., & Spink, A. (2006). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. *Information Processing & Management*, *42*(1), 248–263. https://doi.org/10.1016/j.ipm.2004.10.007

Jansen, B. J., Spink, A., Blakely, C., & Koshman, S. (2007). Defining a session on Web search engines. *Journal of the American Society for Information Science and Technology*, *58*(6), 862–871. https://doi.org/10.1002/asi.20564

Jones, R., Kumar, R., Pang, B., & Tomkins, A. (2007). "I know what you did last summer": Query logs and user privacy. *Proceedings of the Sixteenth ACM Conference on Conference on Information and Knowledge Management*, 909–914. https://doi.org/10.1145/1321440.1321573

Jones, S., Cunningham, S. J., McNab, R., & Boddie, S. (2000). A transaction log analysis of a digital library. *International Journal on Digital Libraries*, *3*(2), 152–169. https://doi.org/10.1007/s007999900022

Kendall, M. G. (1938). A New Measure of Rank Correlation. *Biometrika*, *30*(1/2), 81–93. JSTOR. https://doi.org/10.2307/2332226

Kendall, M. G. (1948). *Rank correlation methods*. Griffin.

Khoo, K., Bolt, P., Babl, F. E., Jury, S., & Goldman, R. D. (2008). Health information seeking by parents in the Internet age. *Journal of Paediatrics and Child Health*, *44*(7–8), 419–423. https://doi.org/10.1111/j.1440-1754.2008.01322.x

Kim, H., & Seo, J. (2006). High-performance FAQ retrieval using an automatic clustering method of query logs. *Information Processing & Management*, *42*(3), 650–661. https://doi.org/10.1016/j.ipm.2005.04.002

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. Journal of Chiropractic Medicine, 15(2), 155–163.

Krogh, A., Brown, M., Mian, I. S., Sjolander, K., & Haussler, D. (1994). Hidden Markov models in computational biology. Applications to protein modeling. *Journal of Molecular Biology*, *235*(5), 1501–1531.

Kurth, M. (1993). The limits and limitations of transaction log analysis. *Library Hi Tech*, *11*(2), 98–104. https://doi.org/10.1108/eb047888

Langville, A. N., & Meyer, C. D. (2011). *Google's PageRank and beyond: The science of search engine rankings*. Princeton university press.

Lee, Y.-S., & Yen, S.-J. (2008). Incremental and interactive mining of web traversal patterns. *Information Sciences*, *178*(2), 287–306. https://doi.org/10.1016/j.ins.2007.08.020

Leskovec, J., Rajaraman, A., & Ullman, J. D. (2020). *Mining of Massive Data Sets*. Cambridge University Press.

Luo, W., & Najdawi, M. (2004). Trust-building measures: A review of consumer health portals. *Communications of the ACM*, *47*(1), 108–113. https://doi.org/10.1145/962081.962089

Mayer, D. K., Terrin, N. C., Menon, U., Kreps, G. L., McCance, K., Parsons, S. K., & Mooney, K. H. (2007). Health Behaviors in Cancer Survivors. *Oncology Nursing Forum*, *34*(3), 643–651. https://doi.org/10.1188/07.ONF.643-651

McNab, C. (2009). What social media offers to health professionals and citizens. *Bulletin of the World Health Organization*, *87*, 566–566. https://doi.org/10.1590/S0042-96862009000800002

Miller, D. R., Leek, T., & Schwartz, R. M. (1999). A hidden Markov model information retrieval system. *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 214–221.

Mohamed, M., & Gader, P. (1996). Handwritten word recognition using segmentation-free hidden Markov modeling and segmentation-based dynamic programming techniques. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *18*(5), 548–554. https://doi.org/10.1109/34.494644

Moorhead, S. A., Hazlett, D. E., Harrison, L., Carroll, J. K., Irwin, A., & Hoving, C. (2013). A New Dimension of Health Care: Systematic Review of the Uses, Benefits, and Limitations of Social Media for Health Communication. *Journal of Medical Internet Research*, *15*(4), e85. https://doi.org/10.2196/jmir.1933

Morahan-Martin, J. (2008). Internet abuse: Emerging trends and lingering questions. In A. Barak (Ed.), Psychological aspects of Cyberspace. Theory, research and applications (pp. 32–69). Cambridge, UK: Cambridge University Press.

Murray, G. C., Lin, J., & Chowdhury, A. (2006). Identification of user sessions with hierarchical agglomerative clustering. *Proceedings of the American Society for Information Science and Technology*, *43*(1), 1–9. https://doi.org/10.1002/meet.14504301312

Navarro-Arribas, G., Torra, V., Erola, A., & Castellà-Roca, J. (2012). User k-anonymity for privacy preserving data mining of query logs. *Information Processing & Management*, *48*(3), 476–487. https://doi.org/10.1016/j.ipm.2011.01.004

Oh, S., Zhang, Y., & Park, M.S. (2016). Cancer information seeking in social question and answer services: Identifying health-related topics in cancer questions on Yahoo! Answers. *Information Research*, *21*(3).

Ozmutlu, H. C., Spink, A., & Ozmutlu, S. (2002). Analysis of large data logs: An application of Poisson sampling on excite web queries. *Information Processing & Management*, *38*(4), 473–490. https://doi.org/10.1016/S0306-4573(01)00043-7

Ozmutlu, S., Spink, A., & Ozmutlu, H. C. (2003). Multimedia Web searching trends: 1997–2001. *Information Processing & Management*, *39*(4), 611–621. https://doi.org/10.1016/S0306-4573(02)00038-9

Pan, W., Shen, C., & Feng, B. (2017). You Get What You Give: Understanding Reply Reciprocity and Social Capital in Online Health Support Forums. *Journal of Health Communication*, *22*(1), 45–52. https://doi.org/10.1080/10810730.2016.1250845

Park, H., & Park, M. S. (2014). Cancer Information-Seeking Behaviors and Information Needs Among Korean Americans in the Online Community. *Journal of Community Health*, *39*(2), 213–220. https://doi.org/10.1007/s10900-013-9784-8

Pazzani, M. J., & Billsus, D. (2007). Content-Based Recommendation Systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *The Adaptive Web: Methods and Strategies of Web Personalization* (pp. 325–341). Springer. https://doi.org/10.1007/978-3-540-72079-9_10

Peker, S., & Koçyiğit, A. (2016). An Adjusted Recommendation List Size Approach for Users' Multiple Item Preferences (Vol. 9883, p. 319). https://doi.org/10.1007/978-3-319-44748-3_30

Perkowitz, M., & Etzioni, O. (1997). Adaptive sites: Automatically learning from user access patterns. *Proc. 6th Int. World Wide Web Conf., Santa Clara, California*, *149*, 158.

Pitkow, J., & Pirolli, P. (n.d.). *MINING LONGEST REPEATING SUBSEQUENCES TO PREDICT WORLD WIDE WEB SURFING*. 13.

Polyzou, A., Nikolakopoulos, A. N., & Karypis, G. (2019). Scholars Walk: A Markov Chain Framework for Course Recommendation. In International Educational Data Mining

Society. International Educational Data Mining Society. https://eric.ed.gov/?id=ED599254

Qiu, L. (1993). Markov models of search state patterns in a hypertext information retrieval system. *Journal of the American Society for Information Science, 44*(7), 413–427.

Quah, D. (1992). *Empirical cross-section dynamics in economic growth*.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257–286. https://doi.org/10.1109/5.18626

Rieh, S. Y., & Xie, H. (Iris). (n.d.). *Patterns and Sequences of Multiple Query Reformulations in Web Searching: A Preliminary Study*. 10.

Rodgers, S., & Chen, Q. (2005). Internet Community Group Participation: Psychosocial Benefits for Women with Breast Cancer. *Journal of Computer-Mediated Communication*, *10*(4). https://doi.org/10.1111/j.1083-6101.2005.tb00268.x

Rose, D. E., & Levinson, D. (2004). Understanding user goals in web search. *Proceedings of the 13th International Conference on World Wide Web*, 13–19. https://doi.org/10.1145/988672.988675

Saha Roy, R., Agarwal, S., Ganguly, N., & Choudhury, M. (2016). Syntactic complexity of Web search queries through the lenses of language models, networks and users. *Information Processing & Management*, *52*(5), 923–948. https://doi.org/10.1016/j.ipm.2016.04.002

Sarukkai, R. R. (2000). Link prediction and path analysis using Markov chains. *Computer Networks*, *33*(1–6), 377–386.

Scanfeld, D., Scanfeld, V., & Larson, E. L. (2010). Dissemination of health information through social networks: Twitter and antibiotics. *American Journal of Infection Control*, *38*(3), 182–188. https://doi.org/10.1016/j.ajic.2009.11.004

Sejal, D., Rashmi, V., Venugopal, K. R., Iyengar, S. S., & Patnaik, L. M. (2016). Image recommendation based on keyword relevance using absorbing Markov chain and image features. International Journal of Multimedia Information Retrieval, 5(3), 185–199. https://doi.org/10.1007/s13735-016-0104-9

Sen, R., & Hansen, M. H. (2003). Predicting Web users' next access based on log data. *Journal of Computational and Graphical Statistics*, *12*(1), 143–155.

Seymore, K., McCallum, A., & Rosenfeld, R. (1999). Learning hidden Markov model structure for information extraction. *AAAI-99 Workshop on Machine Learning for Information Extraction*, 37–42.

Shah, C., Oh, S., & Oh, J. S. (2009). Research agenda for social Q&A. *Library & Information Science Research*, *31*(4), 205–209. https://doi.org/10.1016/j.lisr.2009.07.006

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, *27*(3), 379–423.

Shaw, R. J., & Johnson, C. M. (2011). Health Information Seeking and Social Media Use on the Internet among People with Diabetes. *Online Journal of Public Health Informatics, 3*(1). https://doi.org/10.5210/ojphi.v3i1.3561

Shi, X., & Yang, C. C. (2007). Mining related queries from Web search engine query logs using an improved association rule mining model. *Journal of the American Society for Information Science and Technology*, *58*(12), 1871–1883. https://doi.org/10.1002/asi.20632

Silverstein, C., Marais, H., Henzinger, M., & Moricz, M. (1999). Analysis of a very large web search engine query log. *ACM SIGIR Forum*, *33*(1), 6–12. https://doi.org/10.1145/331403.331405

Spears, W. M. (1998). A compression algorithm for probability transition matrices. *SIAM Journal on Matrix Analysis and Applications*, *20*(1), 60–77.

Spink, A., Jansen, B. J., Wolfram, D., & Saracevic, T. (2002). From e-sex to e-commerce: Web search changes. *Computer*, *35*(3), 107–109. https://doi.org/10.1109/2.989940

Spink, Amanda, & Jansen, B. J. (2006). *Web Search: Public Searching of the Web*. Springer Science & Business Media.

Spink, Amanda, Jansen, B. J., & Cenk Ozmultu, H. (2000). Use of query reformulation and relevance feedback by Excite users. *Internet Research*, *10*(4), 317–328. https://doi.org/10.1108/10662240010342621

Spink, Amanda, Park, M., Jansen, B. J., & Pedersen, J. (2006). Multitasking during Web search sessions. *Information Processing & Management*, *42*(1), 264–275. https://doi.org/10.1016/j.ipm.2004.10.004

Spink, Amanda, Wolfram, D., Jansen, M. B. J., & Saracevic, T. (2001). Searching the web: The public and their queries. *Journal of the American Society for Information Science and Technology*, *52*(3), 226–234. https://doi.org/10.1002/1097-4571(2000)9999:9999<::AID-ASI1591>3.0.CO;2-R

Srikant, R., & Yang, Y. (2001). Mining web logs to improve website organization. *Proceedings of the 10th International Conference on World Wide Web*, 430–437. https://doi.org/10.1145/371920.372097

Tague-Sutcliffe, J. (1992). An introduction to informetrics. *Information Processing & Management*, *28*(1), 1–3. https://doi.org/10.1016/0306-4573(92)90087-G

Tennant, B., Stellefson, M., Dodd, V., Chaney, B., Chaney, D., Paige, S., & Alber, J. (2015). EHealth Literacy and Web 2.0 Health Information Seeking Behaviors Among Baby

Boomers and Older Adults. *Journal of Medical Internet Research*, *17*(3), e70. https://doi.org/10.2196/jmir.3992

Thackeray, R., Neiger, B. L., Burton, S. H., & Thackeray, C. R. (2013). Analysis of the Purpose of State Health Departments' Tweets: Information Sharing, Engagement, and Action. *Journal of Medical Internet Research*, *15*(11), e255. https://doi.org/10.2196/jmir.3002

Tjondronegoro, D., Spink, A., & Jansen, B. J. (2009). A study and comparison of multimedia Web searching: 1997–2006. *Journal of the American Society for Information Science and Technology*, *60*(9), 1756–1768. https://doi.org/10.1002/asi.21094

Turban, E., Sharda, R., & Delen, D. (2010). *Decision Support and Business Intelligence Systems* (9th ed.). Prentice Hall Press.

White, M., & Dorman, S. M. (2001). Receiving social support online: Implications for health education. *Health Education Research*, *16*(6), 693–707. https://doi.org/10.1093/her/16.6.693

Wilson, C. S. (1999). Informetrics. *Annual Review of Information Science and Technology (ARIST)*, *34*, 107–247.

Witten, I. H., Bainbridge, D., & Nichols, D. M. (2009). *How to Build a Digital Library*. Morgan Kaufmann.

Wolfram, D. (2008). Search characteristics in different types of Web-based IR environments: Are they the same? *Information Processing & Management*, *44*(3), 1279–1292. https://doi.org/10.1016/j.ipm.2007.07.010

Wolfram, D., Wang, P., & Zhang, J. (2009). Identifying Web search session patterns using cluster analysis: A comparison of three search environments. *Journal of the American*

Society for Information Science and Technology, *60*(5), 896–910. https://doi.org/10.1002/asi.21034

Xie, H. I. (2008). Users' evaluation of digital libraries (DLs): Their uses, their criteria, and their assessment. *Information Processing & Management*, *44*(3), 1346–1373. https://doi.org/10.1016/j.ipm.2007.10.003

Xie, H. (Iris). (2000). Shifts of interactive intentions and information-seeking strategies in interactive information retrieval. *Journal of the American Society for Information Science*, *51*(9), 841–857. https://doi.org/10.1002/(SICI)1097-4571(2000)51:9<841::AID-ASI70>3.0.CO;2-0

Xie, I., & Joo, S. (2010). Transitions in search tactics during the Web-based search process. *Journal of the American Society for Information Science and Technology*, *61*(11), 2188–2205.

Yang, H.-C., & Lee, C.-H. (2004). A text mining approach on automatic generation of web directories and hierarchies. *Expert Systems with Applications*, *27*(4), 645–663. https://doi.org/10.1016/j.eswa.2004.06.009

Yang, Q., Zhang, H. H., & Li, T. (2001). Mining web logs for prediction models in WWW caching and prefetching. *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 473–478. https://doi.org/10.1145/502512.502584

Yi, K., & Beheshti, J. (2009). A hidden Markov model-based text classification of medical documents. *Journal of Information Science*, *35*(1), 67–81.

Zhang, D., Zambrowicz, C., Zhou, H., & Roderer, N. K. (2004). User information-seeking behavior in a medical Web portal environment: A preliminary study. *Journal of the*

American Society for Information Science and Technology, *55*(8), 670–684. https://doi.org/10.1002/asi.20001

Zhang, J., An, L., Tang, T., & Hong, Y. (2009). Visual health subject directory analysis based on users' traversal activities. *Journal of the American Society for Information Science and Technology*, *60*(10), 1977–1994. https://doi.org/10.1002/asi.21153

Zhang, J., Wolfram, D., Wang, P., Hong, Y., & Gillis, R. (2008). Visualization of health-subject analysis based on query term co-occurrences. *Journal of the American Society for Information Science and Technology*, *59*(12), 1933–1947. https://doi.org/10.1002/asi.20911

Zhang, J., & Zhao, Y. (2013). A user term visualization analysis based on a social question and answer log. *Information Processing & Management*, *49*(5), 1019–1048. https://doi.org/10.1016/j.ipm.2013.04.003

Zhang, J., & Zhao, Y. (2014). Visual Data Mining in a Q&A Based Social Media Website. In C. Chen & R. Larsen (Eds.), *Library and Information Sciences: Trends and Research* (pp. 41–55). Springer. https://doi.org/10.1007/978-3-642-54812-3_5

Zhang, J., Zhao, Y., & Dimitroff, A. (2014). A study on health care consumers' diabetes term usage across identified categories. *Aslib Journal of Information Management*, *66*(4), 443–463. https://doi.org/10.1108/AJIM-01-2014-0008

Zhang, Y., He, D., & Sang, Y. (2013). Facebook as a Platform for Health Information and Communication: A Case Study of a Diabetes Group. *Journal of Medical Systems*, *37*(3), 9942. https://doi.org/10.1007/s10916-013-9942-7

Zhou, B., Hui, S. C., & Fong, A. C. M. (2006). Efficient sequential access pattern mining for web recommendations. *International Journal of Knowledge-Based and Intelligent Engineering Systems*, *10*(2), 155–168. https://doi.org/10.3233/KES-2006-10205

Zhu, J., Hong, J., & Hughes, J. G. (2002). Using markov chains for link prediction in adaptive web sites. *Int. Conference on Soft Issues in the Design, Development, and Operation of Computing Systems*, 60–73.

Zimmer, M. (2013). Patron Privacy in the “2.0” Era: Avoiding the Faustian Bargain of Library 2.0. *Journal of Information Ethics*, *22*(1), 44–59.

Appendix A: The recommendation lists at the topic level

Anchor Topic:	**Recommendation List:**
Cancer	Women's Health
	Men's Health
	Preventive Medicine
	Cardiac/Heart
	Digestive System
	Drugs/Medications
	Senior Health
	Skin/Dermatology
	Brain/Nervous System
	Ears/Hearing
Anchor Topic:	Recommendation List:
Infection	Drugs/Medications
	Women's Health
	Cancer
	Diabetes
	Wellness/Lifestyle
	Nutrition/Food
	Preventive Medicine
	Fitness/Weight Management
	Blood/Blood Pressure
	Cholesterol
Anchor Topic:	Recommendation List:
Heart	Drugs/Medications
	Digestive System
	Public Health
	Children's Health
	Nutrition/Food
	Preventive Medicine
	Respiratory
	Travel Medicine
	Environmental Health
	Vaccines/Immunizations

Appendix B: The recommendation lists for the article level

Anchor Article:	**Recommendation List:**
Learning from Cancer Survivors	Breast Cancer Treatments Offer Patients Many Options
	Older Breast Cancer Survivors Less Likely to Have Follow-up Mammograms
	More Than Survival
	Friendship - and Silliness - Play a Role in Cancer Survival
	Liver Cancer Treatments Offer Longer
	Exploring Alternative Approaches to Cancer Treatment
	Increasing the Survival Rate for Biliary Cancer Patients
	Life after Cancer: Late Effects on Long-Term Survivors
	Quality of Life for Cancer Survivors
	More Than 90% Will Survive Testicular Cancer
	Breast Self-Examination Key to Surviving Cancer
	Study Shows Cryosurgery Extends Options for Liver Tumor Treatment and Improves Survival Rates in Patients
	Opportunities for Childbearing after Cancer Treatment
	The Gold Standard: Research-Based Treatment
	Treatments Save Lives When Skin Cancer Is Detected Early
Anchor Article:	**Recommendation List:**
Individual and Marital Adjustment in Married Couples with Breast Cancer	Two Routes to Breast Cancer Detection
	Breast Cancer Treatments Offer Patients Many Options
	Nutrition and Exercise Tips to Reduce Breast Cancer Risk
	Breast Cancer and Lymph Nodes
	General Information about Breast Cancer
	New Treatment for Early Stage Breast Cancer Reduces Radiation Therapy Duration by 85%
	Reducing Breast Cancer Risks
	New Surgical Procedure for Breast Cancer Patients
	Study Shows Male Breast Cancer Treatment Survival is the Same as in Women
	Communication in Couples with Breast Cancer
	High Volume Breast Cancer Surgeons Have Better Outcomes
	Travel Distance Might Affect Breast Cancer Treatment
	The Dangers of Breast Cancer
	Risk Factors for Breast Cancer
	Breast Cancer Treatment and Diagnosis Update
Anchor Article:	**Recommendation List:**
Breast Lumps and Other Changes	Two Routes to Breast Cancer Detection
	Breast Cancer Treatments Offer Patients Many Options
	Nutrition and Exercise Tips to Reduce Breast Cancer Risk
	Breast Cancer and Lymph Nodes
	General Information about Breast Cancer
	Reducing Breast Cancer Risks
	New Treatment for Early Stage Breast Cancer Reduces Radiation

	Therapy Duration by 85%
	Breast Lumps and Mammograms
	New Surgical Procedure for Breast Cancer Patients
	Study Shows Male Breast Cancer Treatment Survival is the Same as in Women
	Communication in Couples with Breast Cancer
	High Volume Breast Cancer Surgeons Have Better Outcomes
	The Dangers of Breast Cancer
	Risk Factors for Breast Cancer
	Travel Distance Might Affect Breast Cancer Treatment
Anchor Article:	**Recommendation List:**
Nutrition and Exercise Tips to Reduce Breast Cancer Risk	Two Routes to Breast Cancer Detection
	Nutrition and Exercise Tips to Reduce Breast Cancer Risk
	Reducing Breast Cancer Risks
	New Treatment for Early Stage Breast Cancer Reduces Radiation Therapy Duration by 85%
	New Surgical Procedure for Breast Cancer Patients
	Breast Cancer Treatments Offer Patients Many Options
	Prostate Cancer Surgery: Weighing the Risks and Benefits
	High Volume Breast Cancer Surgeons Have Better Outcomes
	Breast Cancer and Lymph Nodes
	Reducing Breast Cancer Risks
	Risk Factors for Breast Cancer
	The Dangers of Breast Cancer
	Travel Distance Might Affect Breast Cancer Treatment
	Nutrition During Childhood Cancer
	Tamoxifen and Breast Cancer
Anchor Article:	**Recommendation List:**
Cancer Genetics: Finding and Treating Cancer Before It Occurs	Breast Cancer Treatments Offer Patients Many Options
	Genetic Testing Offers Information and Dilemmas
	Genetic Screening for Breast Cancer Provides Answers
	Liver Cancer Treatments Offer Longer
	Exploring Alternative Approaches to Cancer Treatment
	Treatment for Cancer of the Larynx
	New and Alternative Treatments for Childhood Cancers
	Breast Cancer Treatment and Diagnosis Update
	The Gold Standard: Research-Based Treatment
	Gene Therapy for Bleeding Disorder Has Potential For Cancer Treatment
	Early Lung Cancer May One Day Be Detected Using Genetic Approach
	Cancer Genetics Screening Program
	Gene Therapy for Patients with Cancers of the Blood
	Opportunities for Childbearing after Cancer Treatment
	Study Shows Cryosurgery Extends Options for Liver Tumor Treatment and Improves Survival Rates in Patients
Anchor Article:	**Recommendation List:**

American Heart Association Dietary Guidelines	Fat and Cholesterol: Planning a Healthy Diet
	New Cholesterol Guidelines for High-Risk Patients
	Using Diet to Lower Your Blood Pressure
	Garlic and Cholesterol
	Statin Drugs Lower Cholesterol
	Want to Reduce Your Cholesterol? These Foods Will Help
	Gelatin Intake and Cholesterol
	Figuring Out Fat
	How to Prevent High Blood Pressure
	Take New Information About Cholesterol with a Grain of Salt
	Managing Cholesterol Levels for Women
	When It Comes to Fats
	Types of Cholesterol
	Guidelines on Use of CPR from the American Heart Association
	Study Says Higher Statin Doses = Lower Cholesterol
Anchor Article:	**Recommendation List:**
Heart Failure Treatable with Careful Diagnosis	Managing High Blood Pressure in Active People
	Does High Blood Pressure Signal Coronary Disease?
	How to Prevent High Blood Pressure
	High Blood Pressure and Kidney Disease
	Chest Discomfort While Walking Could be a Warning Sign
	Using Diet to Lower Your Blood Pressure
	Chest Pain: Wondering
	Heart Failure
	Heart Disease Can Be a Silent Killer
	How Is Coronary Heart Disease Treated?
	Does High Blood Pressure Signal Coronary Disease?
	Defibrillator/Pacemaker Advances Treatment of Heart Failure
	Garlic and Cholesterol
	Pumping Iron Improves Heart Health
	What Causes Heart Failure?
Anchor Article:	**Recommendation List:**
Smoking Before Surgery a Dangerous Decision	Cigarette Smoking Among High School Students Appears to be Declining
	Smoking Teens: Some Good News and Some Bad
	Cigarette Smoking and Birth Control Pills
	Smoke
	What You Can Do to Prevent Teen Smoking
	Cigarette Smoking among Adults in the United States
	Symptoms May Occur After You Quit Smoking
	Three-Pronged Approach Helps Smokers Quit for Good
	Medical College Clinic Turns Smokers Into Quitters
	Alcohol Impairs Tumor-Suppressing Gene in Smokers
	One More Reason Not to Smoke
	For Smokers, Habit Permeates Life
	Smoking and Your Bones
	Link Between Smoking and Drinking -- An Update

	New Report Expands List of Diseases Caused by Smoking
Anchor Article:	**Recommendation List:**
Using Diet to Lower Your Blood Pressure	Managing High Blood Pressure in Active People
	Fat and Cholesterol: Planning a Healthy Diet
	Guidelines on Use of CPR from the American Heart Association
	Take New Information About Cholesterol with a Grain of Salt
	Figuring Out Fat
	Does High Blood Pressure Signal Coronary Disease?
	High Blood Pressure and Kidney Disease
	Salt: It's Everywhere
	Body's Appetite Suppressant Leptin Linked to High Blood Pressure in African Americans
	When It Comes to Fats
	Diet, Exercise Can Stop Diabetes Before It Starts
	Dietary Tips after Gastroenteritis
	Using Diet to Lower Your Blood Pressure
	Moderate Reduction of Dietary Fat has Benefits
	Pregnancy and Calcium in Diet
Anchor Article:	**Recommendation List:**
Salt: It's Everywhere	Managing High Blood Pressure in Active People
	American Heart Association Dietary Guidelines
	Fat and Cholesterol: Planning a Healthy Diet
	Take New Information About Cholesterol with a Grain of Salt
	Guidelines on Use of CPR from the American Heart Association
	Using Diet to Lower Your Blood Pressure
	Figuring Out Fat
	Does High Blood Pressure Signal Coronary Disease?
	High Blood Pressure and Kidney Disease
	Gene May Explain African Americans' Extra Sensitivity to Salt, Leading to High Blood Pressure
	Body's Appetite Suppressant Leptin Linked to High Blood Pressure in African Americans
	Salt and Blood Pressure
	Using Diet to Lower Your Blood Pressure
	Moderate Reduction of Dietary Fat has Benefits
	When It Comes to Fats
Anchor Article:	**Recommendation List:**
Getting Rid of Hepatitis B in the United States	Preventing Hepatitis A Infection While Traveling
	Hepatitis C Treatment is Slowly Improving the Odds
	Hepatitis D
	The Facts about Hepatitis A
	The Facts about Hepatitis C
	Hepatitis B
	Hepatitis Support Group
	Hepatitis C Support Group
	Hepatitis C - Or Its Treatment - Can Cause Fatigue
	Hepatitis B Virus Transmission

	Hepatitis C Virus a Leading Cause of Chronic Liver Disease
	Hepatitis B Can Have Serious Long-Term Consequences
	Handwashing and Vaccines Reduce Incidence of Hepatitis A
	Hepatitis C Virus Infection in Cocaine Users: A Silent Epidemic
	Weight Loss Improves Liver Function in Steatohepatitis
Anchor Article:	**Recommendation List:**
Avian Influenza on the Move	Flu Season Begins in the Southwest
	CDC Addresses Flu Vaccine Shortage
	Flu Season Begins:
	Flu Facts
	How Do I Know It's the Flu?
	One Too Many Flu Shots?
	Season Has Started, But Flu Shot Still a Good Idea
	The Facts about Influenza
	Kicking the Flu
	The Asthma-Influenza Connection
	Vaccinate Now to Prevent Flu Misery Later
	Influenza (Flu)
	Prepare for Flu Season Rather Than Anthrax
	To Prevent the Flu, Schedule Vaccine Now
	Flu Activity Increasing Across Midwest, US
Anchor Article:	**Recommendation List:**
The Facts about HIV Infection and AIDS	Chronic Hives
	Understanding HIV
	Commonly asked Questions about HIV/AIDS Counseling and Testing
	HIV Prevention Program Receives $2.2 Million
	Behavioral Interventions Reduce HIV-Related Sexual Risk Behavior
	Taking New Research in HIV Prevention to the Front Lines
	Transferring Research Results to the Front Line Providers in HIV Prevention
	Should You Seek HIV Counseling and Testing?
	Prevention and Treatment of Hives
	Helping Russia Reduce the Spread of HIV and AIDS
	The Best Way to Know Whether You Are Infected with HIV
	What You Can Do to Protect Yourself and Others from HIV/AIDS
	Combatting the HIV Epidemic in Russia
	Community-Led HIV Intervention Program Helps Reduce AIDS Risk among Inner-City Women
	Understanding the Results of an HIV/AIDS Test
Anchor Article:	**Recommendation List:**
Flu Season Begins:	Flu Season Begins in the Southwest
	CDC Addresses Flu Vaccine Shortage
	Flu Facts
	How Do I Know It's the Flu?
	One Too Many Flu Shots?
	The Facts about Influenza
	Season Has Started, But Flu Shot Still a Good Idea
	The Asthma-Influenza Connection

	Kicking the Flu
	Vaccinate Now to Prevent Flu Misery Later
	Influenza (Flu)
	Prepare for Flu Season Rather Than Anthrax
	Flu Activity Increasing Across Midwest, US
	To Prevent the Flu, Schedule Vaccine Now
	Best Way to Prevent Flu Is to Wash Hands
Anchor Article:	**Recommendation List:**
Smallpox Vaccine Includes a Dose of Risk	Facts About Anthrax and Smallpox as Bioterrorism Weapons
	Typhoid Fever Vaccines
	Flu Time Again, with a Twist: Vaccine by Nasal Spray
	Smallpox Vaccine Research at the Medical College of Wisconsin
	CDC Addresses Flu Vaccine Shortage
	Varicella (Chickenpox) Vaccine is Underused
	Chicken Pox Vaccine
	New Recommendations for Polio Vaccine
	Prevnar Vaccine Further Reducing Childhood Infections
	Age and the Shingles Vaccine
	To Prevent the Flu, Schedule Vaccine Now
	Recommended Childhood Vaccines - 2008
	Rotavirus Vaccine Prevents Most Cases in Infants and Children
	Recommended Childhood Vaccines - 2007
	Handwashing and Vaccines Reduce Incidence of Hepatitis A

Appendix C: The recommendation lists at the Q&A item level

Anchor Q&A item:	**Recommendation List:**
Breast Cancer and Lymph Nodes	Two Routes to Breast Cancer Detection
	Calcifications Very Rarely a Sign of Early Breast Cancer
	Reducing Breast Cancer Risks
	Breast Cancer Treatments Offer Patients Many Options
	New Surgical Procedure for Breast Cancer Patients
	High Volume Breast Cancer Surgeons Have Better Outcomes
	The Dangers of Breast Cancer
	Reducing Breast Cancer Risks
	Risk Factors for Breast Cancer
	Breast Lumps and Mammograms
	Causes of Death among Women
	Tamoxifen and Breast Cancer
	Breast Cysts Preclude Estrogen Use
	Breast Hardening After Radiation Therapy
	College Researchers Studying Breast Cancer Follow-Up Care
Anchor Q&A item:	**Recommendation List:**
Calcifications Very Rarely a Sign of Early Breast Cancer	Reducing Breast Cancer Risks
	Breast Cancer and Lymph Nodes
	New Surgical Procedure for Breast Cancer Patients
	High Volume Breast Cancer Surgeons Have Better Outcomes
	The Dangers of Breast Cancer
	Reducing Breast Cancer Risks
	Risk Factors for Breast Cancer
	Breast Lumps and Mammograms
	Causes of Death among Women
	Tamoxifen and Breast Cancer
	Breast Cysts Preclude Estrogen Use
	Breast Hardening After Radiation Therapy
	College Researchers Studying Breast Cancer Follow-Up Care
	Breast Cancer Treatments Offer Patients Many Options
	Breast Cancer Study Focuses on Post-Menopausal Women
Anchor Q&A item:	**Recommendation List:**
The Dangers of Breast Cancer	Calcifications Very Rarely a Sign of Early Breast Cancer
	Reducing Breast Cancer Risks
	Breast Cancer and Lymph Nodes
	New Surgical Procedure for Breast Cancer Patients
	High Volume Breast Cancer Surgeons Have Better Outcomes
	Reducing Breast Cancer Risks
	Risk Factors for Breast Cancer
	Breast Lumps and Mammograms
	Causes of Death among Women
	Tamoxifen and Breast Cancer
	Breast Cysts Preclude Estrogen Use
	Breast Hardening After Radiation Therapy

	College Researchers Studying Breast Cancer Follow-Up Care
	Breast Cancer Treatments Offer Patients Many Options
	Breast Cancer Study Focuses on Post-Menopausal Women
Anchor Q&A item:	**Recommendation List:**
The Gamma Knife for Non-Invasive Brain Surgery	Breast Cancer Treatments Offer Patients Many Options
	Liver Cancer Treatments Offer Longer
	Treatment for Liver Tumors
	Killing Cancer with Light
	Brain Tumors in Children
	Cryosurgery Freezes and Kills Liver Tumors
	Study Shows Cryosurgery Extends Options for Liver Tumor Treatment and Improves Survival Rates in Patients
	Brain Mapping Provides Direction for Surgeons
	No Causal Evidence for Mobile Phone - Brain Cancer Link
	Brain Tumors in Children
	Treating Brain Tumors with Light
	Brain Aneurysms
	The Gold Standard: Research-Based Treatment
	Study Shows Cryosurgery Extends Options for Liver Tumor Treatment and Improves Survival Rates in Patients
	Exploring Alternative Approaches to Cancer Treatment
Anchor Q&A item:	**Recommendation List:**
Medical Opinions Differ on Testing for Prostate Cancer	Prostate Cancer PSA Testing Faster
	Policy on Prostate Specific Antigen Testing
	Prostate Cancer Surgery: Weighing the Risks and Benefits
	Prostate Cancer Tests Vary in Specificity
	Prostate Cancer Seldom Provides Warning Signs
	Melanoma Cocktail' Allows Testing for Cancer Cells During Surgery
	Milk and Prostate Cancer?
	Cancer Screening on a Budget
	Prostate Cancer Research is Adding Years to Men's Lives
	Prostate Cancer Test Works Well for Black Men
	Nerve-Sparing Prostate Surgery Improves With New Device
	Bleeding after Radiation for Prostate Cancer
	Prostate Test Can't Gauge Full Range of Cancers
	Screening Tests Give Best Information About Health Risks
	Patient Plays a Key Role in Prostate Cancer Treatment Plan
Anchor Q&A item:	**Recommendation List:**
Statin Drugs Lower Cholesterol	New Cholesterol Guidelines for High-Risk Patients
	With Cholesterol
	Using Diet to Lower Your Blood Pressure
	Want to Reduce Your Cholesterol? These Foods Will Help
	Gelatin Intake and Cholesterol
	Fat and Cholesterol: Planning a Healthy Diet
	Types of Cholesterol
	Garlic and Cholesterol
	Managing Cholesterol Levels for Women

	Take New Information About Cholesterol with a Grain of Salt
	Study Says Higher Statin Doses = Lower Cholesterol
	Cholesterol Particle Test
	Increasing Your Levels of “Good” Cholesterol
	Statins Revolutionize Heart Disease Care
	Want to Reduce Your Cholesterol? These Foods Will Help, Not Hurt
Anchor Q&A item:	**Recommendation List:**
Cholesterol Particle Test	New Cholesterol Guidelines for High-Risk Patients
	With Cholesterol
	Using Diet to Lower Your Blood Pressure
	Want to Reduce Your Cholesterol? These Foods Will Help
	Statin Drugs Lower Cholesterol
	Gelatin Intake and Cholesterol
	Fat and Cholesterol: Planning a Healthy Diet
	Types of Cholesterol
	Garlic and Cholesterol
	Managing Cholesterol Levels for Women
	Increasing Your Levels of “Good” Cholesterol
	Take New Information About Cholesterol with a Grain of Salt
	Study Says Higher Statin Doses = Lower Cholesterol
	Statins Revolutionize Heart Disease Care
	Want to Reduce Your Cholesterol? These Foods Will Help, Not Hurt
Anchor Q&A item:	**Recommendation List:**
CPR Not Always the Answer	MCW Researchers Identify Steps to Improve CPR Survival
	Push Hard and Push Fast, Say New CPR Guidelines
	Simpler CPR Technique Helps Bystanders Save Lives
	Guidelines on Use of CPR from the American Heart Association
	Defibrillator/Pacemaker Advances Treatment of Heart Failure
	Heart Failure
	Congestive Heart Failure in the United States: A New Epidemic
	What Causes Heart Failure?
	Heart Failure Treatments
	Chest Pain: Wondering
	Treatment of Chest Pain Patients Differs by Race
	Chest Discomfort While Walking Could be a Warning Sign
	Heart Failure Treatable with Careful Diagnosis
	Angioplasty Now Standard Care for Heart Attacks
	Triglycerides and Heart Attacks
Anchor Q&A item:	**Recommendation List:**
Is It a Stroke Warning?	Lowering the Risk of Stroke
	Latest Emergency Treatment of Strokes
	Blood-Thinning Drug's Benefit in Stroke Prevention
	Stroke Program Provides Pioneering Treatment
	Understanding and Preventing Subsequent Strokes
	Frequently Asked Questions About Stroke
	Hormone Replacement and Stroke
	Common Condition Emerges as Possible Stroke Cause
	Many Techniques and Disciplines Contribute to Stroke Rehabilitation
	Medical College Stroke Research Could Lead to New Therapies

	The Stroke Club at Froedtert
	Recovery Limited by Severity of Stroke
	Snake Venom Research Drug Can Reduce Stroke Disability
	Froedtert & Medical College Recognized for Stroke Treatment
	Estrogen and the Risk of Stroke
Anchor Q&A item:	**Recommendation List:**
Types of Cholesterol	New Cholesterol Guidelines for High-Risk Patients
	With Cholesterol
	Want to Reduce Your Cholesterol? These Foods Will Help
	Garlic and Cholesterol
	Statin Drugs Lower Cholesterol
	Using Diet to Lower Your Blood Pressure
	Gelatin Intake and Cholesterol
	Fat and Cholesterol: Planning a Healthy Diet
	Study Says Higher Statin Doses = Lower Cholesterol
	Managing Cholesterol Levels for Women
	Take New Information About Cholesterol with a Grain of Salt
	Increasing Your Levels of “Good” Cholesterol
	Cholesterol Particle Test
	Statins Revolutionize Heart Disease Care
	Want to Reduce Your Cholesterol? These Foods Will Help, Not Hurt
Anchor Q&A item:	**Recommendation List:**
Urinary Tract Infections Common in Women, but Treatable	Interstitial Cystitis of the Bladder
	Symptoms and Diagnosis of Urinary Tract Infection (UTI)
	Risks for Urinary Tract Infection (UTI)
	The Cause of Urinary Tract Infections
	Cystocele (Fallen Bladder)
	Treatment of Urinary Tract Infection (UTI)
	Menopause and Bladder Control
	General Information about Bladder Control for Women
	Treatment for Urinary Incontinence in Children
	Kegel Exercises and Bladder Control
	Bladder Control Problems: Medicines May Be the Cause
	Urinary Incontinence in Children
	Causes and Treatments for Bladder Control Problems in Women
	Pregnancy, Childbirth and Bladder Control
	Preventing Bladder Infections
Anchor Q&A item:	**Recommendation List:**
Flu Season Winding Down, But Not Over	Flu Season Begins in the Southwest
	CDC Addresses Flu Vaccine Shortage
	Flu Season Begins:
	Flu Facts
	How Do I Know It’s the Flu?
	One Too Many Flu Shots?
	The Facts about Influenza
	Season Has Started, But Flu Shot Still a Good Idea

	Kicking the Flu
	The Asthma-Influenza Connection
	Vaccinate Now to Prevent Flu Misery Later
	Influenza (Flu)
	Prepare for Flu Season Rather Than Anthrax
	To Prevent the Flu, Schedule Vaccine Now
	Flu Activity Increasing Across Midwest, US
Anchor Q&A item:	**Recommendation List:**
Bladder Infections More Likely in Women	Interstitial Cystitis of the Bladder
	Symptoms and Diagnosis of Urinary Tract Infection (UTI)
	Risks for Urinary Tract Infection (UTI)
	Urinary Tract Infections Common in Women, but Treatable
	The Cause of Urinary Tract Infections
	Treatment of Urinary Tract Infection (UTI)
	Cystocele (Fallen Bladder)
	Treatment for Urinary Incontinence in Children
	General Information about Bladder Control for Women
	Bladder Control Problems: Medicines May Be the Cause
	Kegel Exercises and Bladder Control
	Urinary Incontinence in Children
	Causes and Treatments for Bladder Control Problems in Women
	Menopause and Bladder Control
	Preventing Bladder Infections
Anchor Q&A item:	**Recommendation List:**
Season Has Started, But Flu Shot Still a Good Idea	Flu Season Begins in the Southwest
	Flu Season Begins:
	CDC Addresses Flu Vaccine Shortage
	Flu Season Winding Down, But Not Over
	How Do I Know It's the Flu?
	Flu Facts
	Best Way to Prevent Flu Is to Wash Hands
	One Too Many Flu Shots?
	The Facts about Influenza
	Kicking the Flu
	Vaccinate Now to Prevent Flu Misery Later
	The Asthma-Influenza Connection
	Influenza (Flu)
	Prepare for Flu Season Rather Than Anthrax
	To Prevent the Flu, Schedule Vaccine Now
Anchor Q&A item:	**Recommendation List:**
Shingles Pain Can Last Long After Rash Heals	Shingles: An Explanation
	For Shingles, It's Treatment ASAP
	Genital Herpes
	The Facts about Shingles
	Chickenpox
	Genital Herpes (Herpes Simplex Virus)

	Coping with the Pain of Shingles
	Immunization Against Shingles
	Shingles Pain May Persist
	Should Your Child Be Vaccinated for Chicken Pox?
	Shingles
	Cold Sores and Chicken Pox
	Vaccine Can Prevent Shingles and Its Debilitating Pain
	Vitamin E and Shingles
	Varicella (Chickenpox) Vaccine is Underused

Curriculum Vita

Xin Cai

School of Information Studies
University of Wisconsin-Milwaukee

EDUCATION

Ph.D. | 2014-2022 (Expected) | University of Wisconsin-Milwaukee

- Major: Information retrieval
- Dissertation: "Application of the Markov chain method in a health portal recommendation system"
- Dissertation Committee: Dr. Jin Zhang (chair), Dr. Dietmar Wolfram, Dr. Iris Xie, Dr. Xiangming Mu, Dr. Kun Lu

M.A. | 2012-2014 | Central China Normal University

- Major: Information Science
- Advisor: Dr. Lixin Xia, Central China Normal University

B.A. | 2007-2011 | Shenyang Normal University

- Major: Computer Science

RESEARCH INTEREST

- Data science
- Statistical analysis and text analysis
- Consumer health informatics
- Information retrieval
- Recommendation system

PUBLICATIONS

Refereed Journal Articles

1 Zhang, J., Zhao, Y., Cai, X., Le, T., Fei, W., & Ma, F. (2020). A Comparison of Retrieval Result Relevance Judgments Between American and Chinese Users. *Journal of Global Information Management (JGIM)* 28(3), 148-168.
2 Zhang, J., Cai, X., Le, T., Fei, W., & Ma, F. (2019). A Study on Effective Measurement of Search Results from Search Engines. *Journal of Global Information Management (JGIM), 27*(1), 196-221.
3 Zhang, J., Wang, Y., Zhao, Y., & Cai, X. (2018). Applications of inferential statistical methods in library and information science. *Data and Information Management*, *2*(2), 103-120.
4 Lu, K., Cai, X., Ajiferuke, I., & Wolfram, D. (2017). Vocabulary size and its effect on topic representation. *Information Processing & Management*, 53(3), 653-665.

5 Smiraglia, R. P., & Cai, X. (2017). Tracking the Evolution of Clustering, Machine Learning, Automatic Indexing and Automatic Classification in Knowledge Organization. *Knowledge Organization*, 44(3), 215-233.

Refereed Conference Posters

1. Zhao, Y., Wang, Y., & Xin, C. (2018). *A citation-based review of study on image retrieval*. iConference 2018, Sheffield, UK.

AWARDS AND FELLOWSHIPS

Distinguished Dissertation Fellowship 2018

· Highly honorable and competitive fellowship offered by Graduate School at the UWM ($16,500)

Chancellor's Graduate Student Awards 2014, 2016-2019

TEACHING EXPERIENCE

Instructor School of Information Studies, University of Wisconsin-Milwaukee

· Undergraduate, 240 Web Design I, Fall 2018-Spring 2021

Teaching Assistant School of Information Studies, University of Wisconsin-Milwaukee

· Undergraduate (Online), 310 Human Factors in Information Seeking and Use, 2017 Fall

· Undergraduate (Onsite), 110 Introduction to Information Science, 2017 Spring

· Graduate (Online), 511 Metadata, 2016 Fall