Karun Bhai Bajracharya

Analysis of User's Data for Book Search

Acknowledgements

During my thesis, a number of people, who supported me with all the assistance they could grant. Their suggestions and comments for my work were appreciable and effective throughout. The people who bestowed me with their profuse help are namely:

- Prof. Nils Pharo, Department of Social Science, HIOA
  (Thesis Supervisor)
- My mom, dad and my family.
- Sushil Chitrakar and Ambika Shrestha
- Stutee Tamrakar

Through the thesis that is presented here, I would like to extend my sincere thanks and gratitude to all those personals who have lent their help directly and indirectly during my thesis. I will always cherish their support, guidance and encouragement that strengthen me to achieve my goal.

Abstract

Today the world is flooding with enormous amount of information in the web. For retrieval of proper information, it can be a confusing task though with help of some effective tools we can seek desired information within less amount of time. While searching for a book, the users are confronted with both professional metadata and user-generated content. The goal of my thesis is to analyze behavior of the user using these two sources for a relevant book search. As there are two type of tasks, goal oriented task and non-goal oriented task, depending on these tasks the user activity was analyzed. Also, two system interface design was observed on how the user performances are accelerated with multiple search functions.

Table of contents

Chapter 1: Introduction

1.1 Topic

Book search means search of relevant document in regard of necessity and requirement of user in a short period of time. Analysis of the user's behavior is essential part of search activity. The search is basically based on effective process carried out using various accurate keywords. This thesis focuses on analyzing the data of users in order to effectively and efficiently search books from a data set.

1.2 Problem description and motivation

As there is huge information online, main problem is to locate relevant and accurate information which is a tedious task in itself. Firstly, to identify proper and effective keyword is basic necessity. Secondly, repetition of identical information in different categories can be problematic. For example, if a user needs a specific book but there are numerous alternatives of same book in different topics and sites, this can be inconvenient. Also, quick and accurate retrieval of information is essential with the increase of data or documents each day, search of related information may get hectic. Thus, this thesis is motivated by errors in everyday book search and desire for convenient user facility.

1.3 Research Question

RQ.1. How do users use professional metadata and user-generated content in a book search?

Professional metadata is a detailed product descriptions of book and production date which are created by professionals or publishers whereas user-generated content are contents given by user such as reviews, user-tags and ratings given to specific book (Gäde et al, 2015). This research question seeks to find out some of the important professional metadata and user-generated content that most of the users use in a book search.

RQ.2. How do users behave in goal oriented task and non-goal oriented task?

Analysis of user behavior is done in two parts: goal oriented and non-goal oriented (http://social-book-search.humanities.uva.nl/#/interactive). Goal oriented task depends upon performance carried out by the user with adequate set of instructions provided. Whereas non-goal oriented task depends upon performance carried out by participants

themselves with own numerous style of search. This research question seeks to find out behavior of users in goal oriented and non-goal oriented tasks.

RQ.3. Is there any difference found from the system interface design in the user's behavior?

There can be different interfaces which are designed with various book search functionalities and behavior of users may vary with the interface. This research question tries to find out any differences in user's behavior linked with the design of different interfaces.

1.4 Contribution

This thesis would definitely contribute to literature and further study of user's behavior during search of appropriate and precise information. Also, a system would be enhanced and helpful for any user with respect to their search activity.

1.5 Thesis outline

This section provides the overview of contents in each section of this thesis. The chapters and the details contained in it are described as below:

- Chapter 2: In this chapter the literature reviews on previous works related to my topic has been described. Discussions of previous literature on professional metadata, user-generated content and folksonomies are discussed in this section.
- Chapter3: This chapter discusses about the methodology on procedure in collection of data and data analysis methods.
- Chapter 4: This chapter provides the data analysis details on each participant for two tasks done in two interfaces.
- Chapter 5: This chapter presents the summary of findings and discussion of this research work.
- Chapter 6: This chapter provides the limitations and future work that can be done in near future.
- Chapter 7: This chapter discusses on the conclusion made on each of the research questions.

Chapter 2: Literature Review

There had been conducted various experiments relating to information retrieval in the past in initiative for evaluation of XML retrieval (INEX) from the year 2004-2010. The main purpose of the INEX is to investigate effects of some features added to existed information that affects in behavior of user in searching relevant information depending on the tasks, user interface and the system in which they search. (Nordlie & Pharo, 2012).

Some of relevant findings that helped literature review were found in the overview of the INEX 2009 interactive track. In this experiment the users were asked to find relevant information based on tasks given. With the help of more structured data than previous year's data consisting of traditional bibliographic metadata, user-generated tags, promotional texts and reviews from publishers and professional reviewers. In this experiment the user have to rate different kinds of metadata and user-defined content from 1-5. Most ratings were given to reviews and publisher's descriptions as 3.32 and 3.5 respectively from overall data. (Nordlie & Pharo, 2012)

The goal to find relevant information is a very tedious task unless and until there is something more precise and accurate information on the content and these contents can be metadata. The structured information which describes, explains, locates or makes it easier to retrieve, use or manage an information resource is termed as metadata. It is often called data about data or information about information. (NISO press, 2001) So with these means of data, information retrieval can be easily fulfilled and in less amount of time.

Understanding what is professional or formal or standard metadata and what's the purpose for its creation is important to identify. Long term preservation of metadata in a given project primarily depends on associated data and intended repository, which are the standards numerous metadata. It can be standardized by formal and standards which includes information about outputs research in standardized format having controlled vocabulary accepted and used by community. (Strasser, 2015) If some user wants to search information on what he/she wants then they can easily retrieve required and relevant information by going through metadata attached within information.

Content indexing is the method that describes document by means of enrichment done in single word or entire sentence. This helps to speed up or accelerate relevance decision in retrieval of any kind of information. So what can be seen is that for preservation and easy retrieval of

information, formal metadata plays a vital role. This citation from (Bertman, 2005) was taken from the book folksonomies, indexing and retrieval in web 2.0.

Problem solving for retrieval of relevant information to a certain extent is done by folksonomies. So structured metadata provided by third parties helps in retrieving accurate documents of interest that contains descriptions added by the user or some other users. (Hammond et al., 2005).

User-tags, reviews and ratings that are created by the user add another dimension while searching for the specific book. The contents which are created by the user are termed as user-generated content. (Bogers & Petras, 2015)

Chapter 3: Methodology

Qualitative analysis is done with data collection from Oslo and Akershus University College of Applied Sciences (HIOA).

3.1 Data collection:

Collection of 11 data inputs provided by HIOA has been analyzed. These data determine the behavior of users on retrieval of information and were collected for "The CLEF (Conference and labs of the evaluation forum) 2015 Social Book Search (SBS) Lab". The main objective of CLEF is to encourage research, originality and improvement of information access systems with emphasis on multilingual information in various levels of configurations. Following are the citation of SBS (http://www.clef-initiative.eu//):

"The CLEF Initiative is structured in two main parts:

1. a series of Evaluation Labs, i.e. laboratories to conduct evaluation of information access systems and workshops to discuss and pilot innovative evaluation activities;
2. a peer-reviewed Conference on a broad range of issues, including
    - investigation continuing the activities of the Evaluation Labs;
    - experiments using multilingual and multimodal data; in particular, but not only, data resulting from CLEF activities;
    - research in evaluation methodologies and challenges"

In many scenarios of book search, SBS lab basically inquires users search more than an inquiry and seek more professional metadata. Simple search based on recommendation and queries of book is generally focus of the research but real world information has more complexity. The objective is to research and enhance techniques for support to users in intricate book search.

"The document collection consists of 1.5 million book descriptions with metadata from Amazon and Library Thing." This has been extracted from SBS website (http://social-book-search.humanities.uva.nl/#/interactive). In the context of Amazon, professional metadata are book title, publisher, author, publication year, library classification codes, other categories and product information. Also user-generated content are contents in the form of user ratings and

reviews. User tags and user created metadata on book characters, locations, awards, recommendations, advertisements and other details are extracted from Library Thing.

Software 'Morae' was used during recording of data. There are basically two tasks with descriptive instructions on how to be familiar with the system. In both tasks, users were first introduced to a task making them comfortable with the system. Then they were assigned with actual tasks. These tasks were taken from Social Book Search website (http://social-book-search.humanities.uva.nl/#/interactive). In task 1 and 2 contains various tasks and their contents are listed below:

Goal oriented task: Task 1

Imagine you participate in an experiment at a desert-island for a month. There will be no people, no TV, radio or other distraction. The only things you are allowed to take with you are 5 books. Please search for and add 5 books to your book-bag that you would want to read during your stay at the desert-island.

- Select one book about surviving on a desert island.
- Select one book that will teach you something new.
- Select one book about one of your personal hobbies or interests.
- Select one book that is highly recommended by other users (based on user ratings and reviews).
- Select one book for fun.

Please add note (in the book-bag) explaining why you selected each of the five books. So the users have to follow book search based on the task mentioned above.

Non-goal oriented task: Task 2

Imagine you are waiting to meet a friend in a coffee shop or pub or airport or your office. While waiting, you come across this website and explore it looking for any book that you find interesting, or engaging or relevant. Explore anything you wish until you are completely and utterly bored. When you find something interesting, add it to the book-bag. Please add a note (in the book-bag) explaining why you selected each of the books.

The users have to follow the book search on the basis of above task mentioned.

Snapshots of interface 1 and 2:

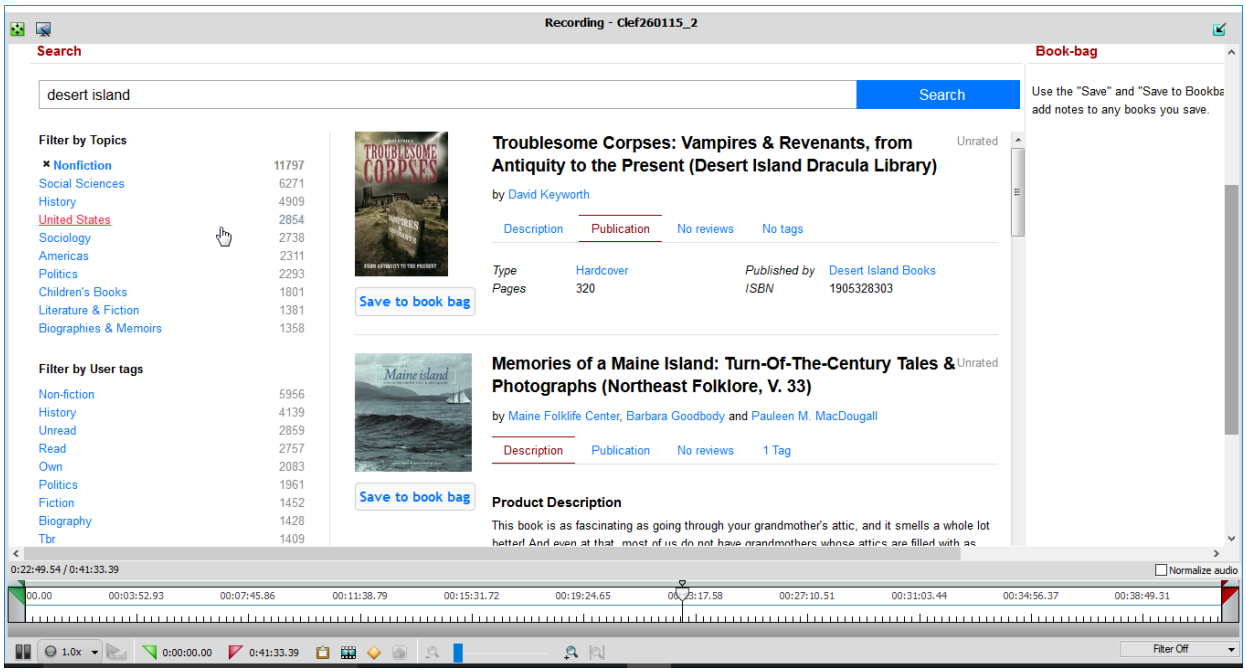


Figure 3.1.a Interface 1

The figure 3.1.a shown above is the interface 1which has search functionality, filter by topics and filter by user tags and book-bag.

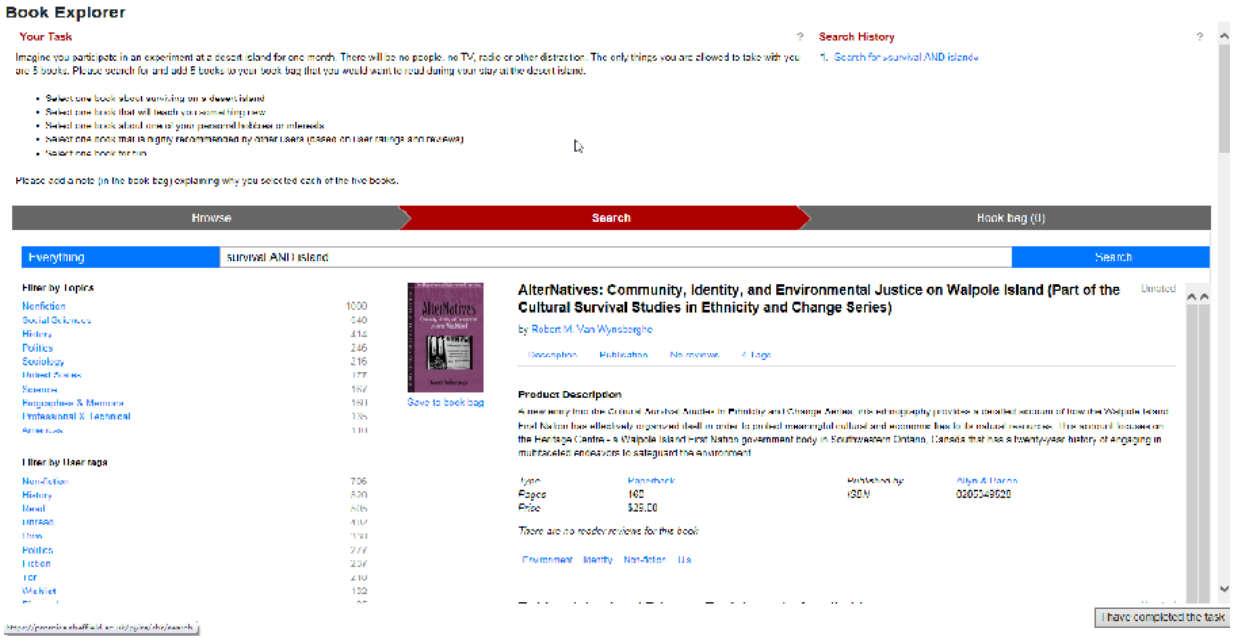


Figure 3.1.b Interface 2.

Above figure 3.1.b shows the interface 2 that actually contains browse function, search functionality and book-bag.

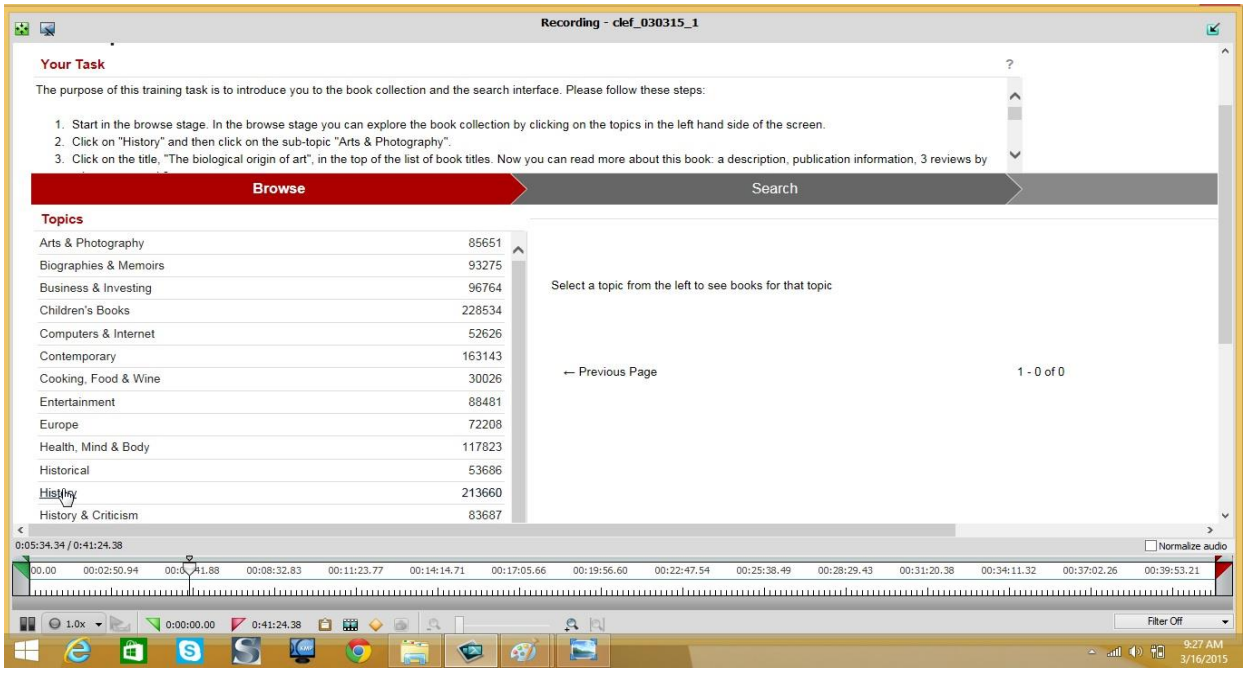


Figure 3.1.c Interface 2

In above figure 3.1.c the browse function can be clearly seen which contains different topics and which makes user easier to explore on the varieties of topics.

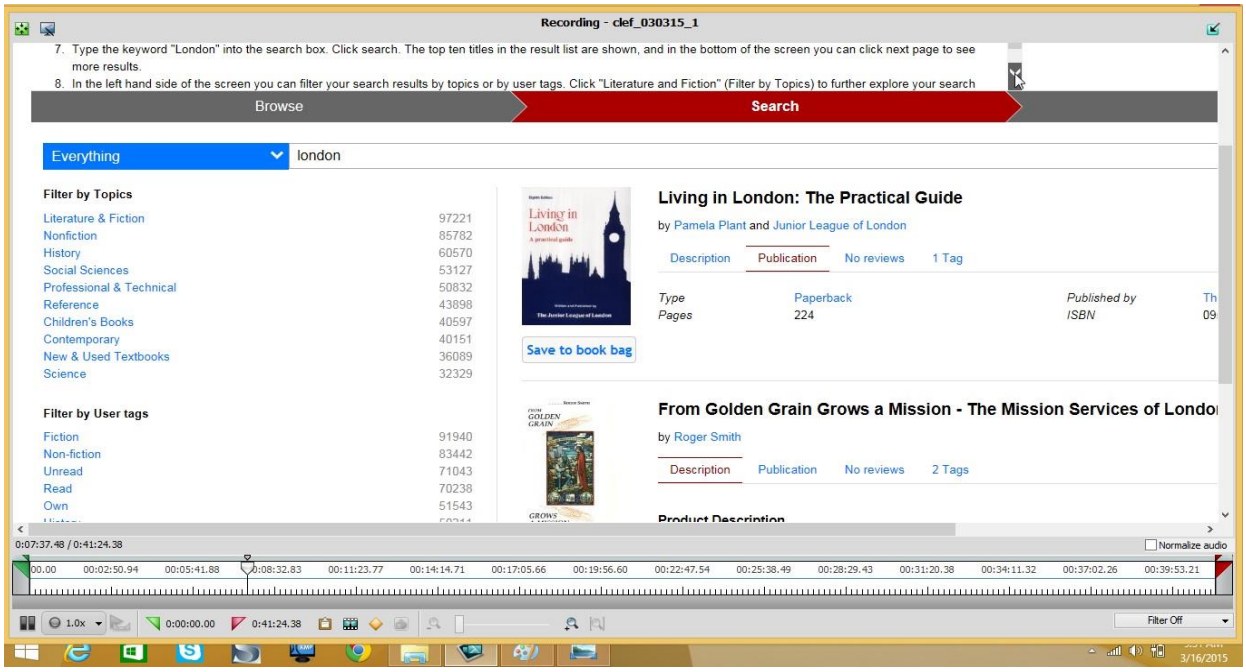


Figure 3.1.d The search functionality in interface 2.

In above figure 3.1.d search functionality can be seen where there are filter by topics and filter by user-tags option.

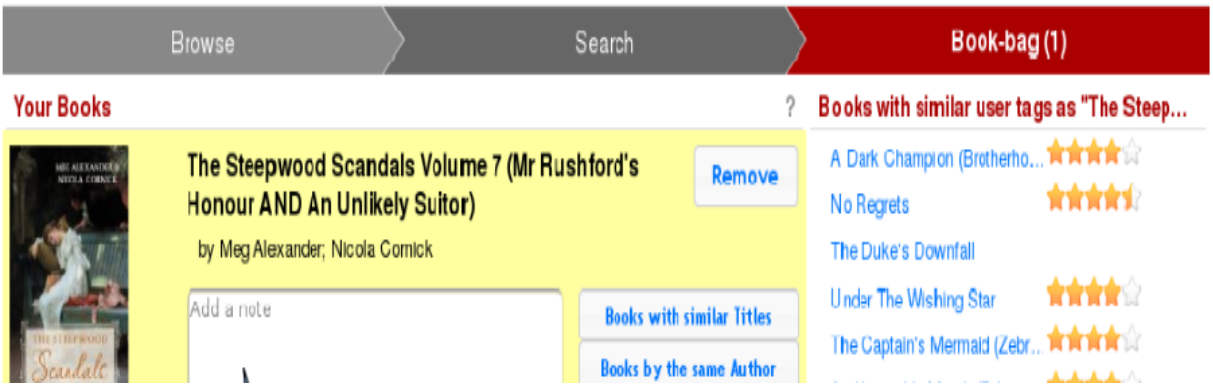


Figure 3.1.e The book-bag section of the interface 2.

Figure 3.1.e shows the book-bag after the specific book has been chosen then it shows up here and it also contains add note box to comment on why that book was chosen. Apart from the other interface in this interface the user can see a list of book under various topics appears in the browse function.

3.2 Data Analysis method:

Data's collected were visually analyzed using Morae Software (Manager Mode) (https://www.techsmith.com/morae-features.html). The main purpose of this analysis is observation of user's behavior during usage of professional metadata and user-generated content.

During book search, two tasks needed to be performed. In case of task 1, Surviving in Desert Island, the user had to choose book according to sub tasks provided. This goal oriented task required user to choose particular book as requested. From their choice of book, I tried to mainly notice how the user focuses on varied options of metadata and user generated content while finalizing the book. Their behavior from beginning till end such as keyword search, description, reviews, publication date and user tags has been perceived to create tables using Morae. The number of books users had gone through while completing other sub tasks are listed in detail in table 4.1.a. of chapter 4 in data analysis section.

In case of task 2, Waiting for a friend, the user had to choose book according to their own preference. This non-goal oriented task uses same strategy as in task 1 for analysis which concentrates on selection of usage of both metadata while confirming the book. The number of

books users had gone through while browsing through diverse books are enlisted in data analysis section in table 4.1.b. of chapter 4.

The role of two types of interface in both tasks were analyzed according to the performance of user during book search with formal and informal metadata however browse function is an extra feature in interface 2.

Chapter 4: Data analysis

The data analysis of 11 participants has been analyzed and illustrated in the table below. Here symbol "Y" stands for choice of the participants. For task 2: 1st, 2nd, 3rd and so on denotes the number of books selected and First, Second, Third and so on denotes books searched.

4.1 Participant no.1

Task 1: Here the interface 1 was used.

Sub-task	Books searched	Descriptions	Reviews	Publication	Tags	Book Bag
A	1	Y	Y	Y		3rd book selected
	2	Y				
	3	Y				
B	1					Selected
C	1	Y				Selected
D	1				Y	3rd book selected
	2	Y	Y	Y		
	3		Y			
E	1		Y			Selected

Table 4.1.a

Sub-task A: The participant inserted query related to content 'survival in Desert Island'. List of books were displayed and the participant's prioritization was first publication date, then reviews and descriptions. The first book was not selected so the user switched on to next two books and studied its description. Thus 3rd book was selected in book bag.

Sub-task B: The participant searched for book that 'teaches something new'. Here the user directly checked for his/her favorite topic without going through detailed metadata.

Sub-task C: The participant entered queries regarding personal 'hobbies or interests'. The description was read thoroughly and book was finalized in book bag.

Sub-task D: For selection of book that was 'recommended books by other users', the participants went through reviews and ratings of various books. In 1st book user tags were

browsed then user changed to 2nd book where firstly description was seen, followed by publication date and then reviews. Finally the user selected 3rd book to book bag after studying reviews.

Sub-task E: The participants started book search 'for fun'. For this reviews were followed and selected in book bag. Thus, the task 1 was carried out with usage of both professional metadata and user generated metadata.

Task 2: Here the interface 1 was used.

No. of books selected	Books searched	Descriptions	Reviews	Publication	Tags	Book Bag
1	1					selected
2	1					selected
3	1	Y	Y			2nd book selected
	2	Y		Y		
4	1					Selected
5	1					selected
6	1					selected
7	1	Y		Y		selected

Table 4.1.b

For this task, there were no limitations in study and selection of books and it depended on the users themselves as it is non-goal oriented task. So, the participants selected a total of seven books on their own will. The selection of 1st book was based on participant's favorite writer so the user typed name in search box and selected the book. Similar activity is carried out in selection of 2nd book. So, in both cases the book was selected based on their prior knowledge.

In the selection procedure of 3rd book, the user entered writers name and then filtered the topics and choose literature. He/she went through description and reviews of a book but did not select it and read description and publication date of another book which was selected. For the 4th book, queries were asked based on title of book and choose novel category under filter by user tags option. List of books were displayed but again the user changed the book category, under filter by topic, to literature. The user continued search with book title and finalized the book in book bag.

In the search of both 5th and 6th book, participants typed in same writers name in search box and choose their desired book without usage of metadata. It was interesting because the candidate had good knowledge on the topic. The 7th book selection was based on study of publication date and descriptive information of book.

4.2 Participant no.2

Task 1: Here interface 1 was used.

Sub-task	Books searched	Descriptions	Reviews	Publication	Tags	Book Bag
A	1	Y				Selected
B	1	Y		Y		Selected
C	1			Y	Y	2nd book selected
	2	Y			Y	
D	1	Y				2nd book selected
	2		Y			
E	1	Y			Y	Selected

Table 4.2.a

Sub-task A: The participant selected non-fiction topic first. Then on the search box ‘survival on desert’ was typed and among numerous books, user selected one and went through its description. Finally book was selected.

Sub-task B: In filter by topic, the user selected contemporary and searched for calligraphy. Among the books displayed, first the user viewed publication date, then description and choose a book.

Sub-task C: First keyword clothes were typed in search box. Then in non-fiction section, contemporary section was further selected where 1st book was observed on the basis of publication date and user tags. For 2nd book, description and tags were followed and selected in book bag.

Sub-task D: Similar topic as in sub-task C from user tag option was used. The description of 1st book was seen but the user changed to 2nd book. This book was finalized after considering the reviews.

Sub-task E: In this task, mystery and thriller were written in search box initially. Then user tags were viewed with description as well. And the book was confirmed.

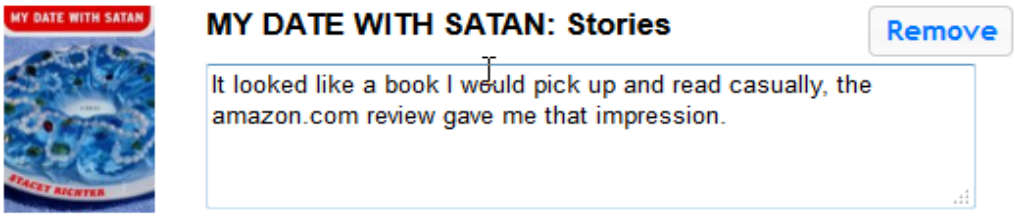


Figure 4.2.a Snapshot of comment by participant no.2.

Task 2: Here interface 1 was used.

No. of books	Books searched	Descriptions	Reviews	Publication	Tags	Book Bag
1	1	Y				Selected
2	1	Y				Selected
3	1	Y				Selected
4	1			Y		Selected
5	1	Y				Selected
6	1					Selected

Table 4.2.b

In this task, the participant had chosen 6 books in total. The 1st, 2nd, 3rd and 5th books were directly selected on the basis of description of books. Whereas in case of 4th book, publication dates were analyzed in detail and then chosen. For 6th book, the user went through various topics and at last chose on basis of writer's name.

4.3 Participant no.3

Task 1: Here interface 2 was used.

Sub-task	Books searched	Descriptions	Reviews	Publication	Tags	Book Bag
A	1	Y	Y	Y		2nd book selected
	2	Y				
B	1	Y				Selected
C	1			Y		Selected
D	1		Y	Y		1st book selected
	2					

	3					
E	1		Y			5th book selected
	2		Y			
	3	Y				
	4	Y				
	5	Y				

Table 4.3.a

Sub-task A: The search was started from non-fiction section and description, publication date and reviews were seen but the user changed to another book. And 2nd book was finalized by reading description of the book.

Sub-task B: The browse function is available in interface 2, so here the participant used historical topic under this function first. Then switched to science topic and selected astronomy. Under this section a book was confirmed in book bag after going through the description.

Sub-task C: The browse function was used in this sub-task as well. First science was selected and sub categories were viewed. Arts and photography was selected and again sub topic of design and decorative was confirmed. But the participant for a second time searched for painting in search box. Among the list of books choose one and book was confirmed after viewing publication date.

Sub-task D: The topic history and philosophy which lies under science was browsed. 1st books description and reviews were seen. The book had 5 star rating. Also, 2nd and 3rd book under same category was viewed but at the end 1st book was selected. Here it can be evaluated that reviews and ratings play vital role in selection of books.

Sub-task E: The same topic as in sub-task D was gone through. The 1st book had 4 star rating reviews in it but the user changed to 2nd book which had 5 star rating. Now the user changed the topic in browse function into science fiction and fantasy. Here 3rd, 4th and 5th books were viewed along with their descriptions. Finally 5th book was selected.

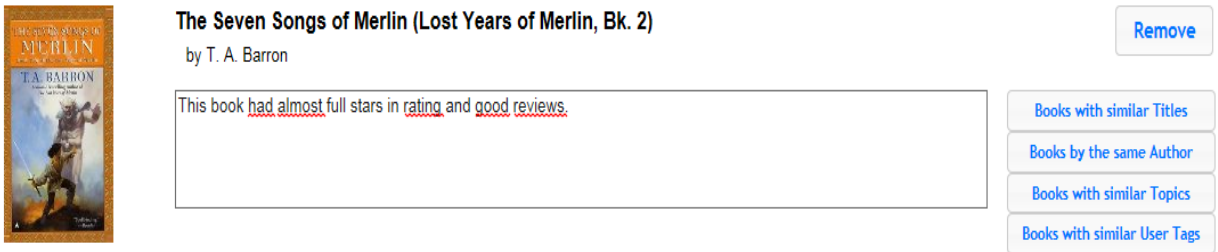


Figure.4.4.a Snap shot of comment by participant no.3.

Task 2: Here interface 2 was used.

No. of books	Books searched	Descriptions	Reviews	Publication	Tags	Book Bag
1	1	Y				Selected
2	1	Y		Y		3rd book selected
	2			Y		
	3			Y		
3	1		Y	Y		Selected
4	1	Y		Y		2nd book selected
	2	Y	Y	Y		
5	1	Y	Y			5th book selected
	2	Y				
	3	Y				
	4	Y				
	5	Y	Y			
6	1	Y				Selected
7	1	Y				3rd book selected
	2		Y			
	3	Y	Y			
8	1	Y				2nd book selected
	2	Y	Y			
9	1	Y				2nd book selected
	2			Y		
10	1	Y				4th book selected
	2		Y			
	3	Y	Y			
	4	Y	Y			
11	1					Selected

Table 4.3.b

In this task, the participant was quiet involved in book selection task because he/she had selected a total of 11 books. Description of the book under topic arts and photography was browsed for confirmation of 1st book. For 2nd book selection, similar procedure as in 1st book was followed but user choose one from two books after looking into publication date and

description of book. For 3rd book, same topic was chosen and selected after short evaluation of production date and reviews. Even for 4th book, same topic was selected but the user went through a sequence of search from description to reviews to publication date between two books and one among two was confirmed in book bag.

Browse function was considered in finalization of 5th book. Here the user went through five books under the category 'computers and internet'. Descriptions of all five books were seen respectively. Both second, third and fifth book, in this task of book selection procedure, were reviewed. The third book was under category of science fiction and fantasy. But the user chose fifth book. For selection of 6th book same category as above was assessed and description of the book was checked. The 7th book was chosen among three books under the same topic general fiction. Only description was seen in first book and only review was seen in second book. But both description and review were viewed in third book and it was selected.

The 8th book was selected among two books with topic literature. The description of first book was checked. Both description and reviews were checked in second book and it was chosen. In 9th book, under category puzzle and games, two books were chosen. Description of first book was seen and for second book publication date was seen. The user selected second book in book bag. In 10th book selection, four books were browsed. Two books were under the same topic as in 9th book and other two were under category StarCraft. Description and reviews were followed in all books and last book was confirmed in book bag. In 11th book, topic game of thrones was searched and maybe because of prior knowledge of book already, the user chose the book without usage of formal metadata and informal metadata.

4.4 Participant no.4

Task 1: Here interface 2 was used.

Sub-task	Books searched	Descriptions	Reviews	Publication	Tags	Book Bag
A	1	Y		Y		4th book selected
	2	Y		Y		
	3	Y				
	4	Y				
B	1	Y			Y	Selected

C	1	Y		Y		2nd book selected
	2			Y		
D	1	Y				Selected
E	1	Y				Selected

Table 4.4.a

Sub-task A: The keyword survival guide was punched in book search and list of books were displayed. For bagging 1st and 2nd book - book bag, description and publication date were reviewed. In case of 3rd book, description was not available so the user shifted to 4th book and after reviewing, selected the book.

Sub-task B: Keyword, harvesting food was typed in search functionality section. Descriptions and tags were gone through respectively and thus book was selected.

Sub-task C: Search engine was utilized to find author and title of a book. For selection of 1st book, publication date and descriptions were reviewed thoroughly. Again for 2nd book selection the publication book was reviewed.

Sub-task D: Same topic as in previous task had been searched. The book description was seen and selected.

Sub-task E: For selection, books from Shakespeare were searched and after description were followed up.

Task 2: Here interface 2 was used.

No. of books	Books searched	Descriptions	Reviews	Publication	Tags	Book Bag
1	1	Y				2nd book selected
	2					
2	1					Selected
3	1					Selected
4	1	Y				2nd book selected
	2	Y				
5	1					Selected
6	1					Selected
7	1					Selected

8	1					Selected
9	1					Selected

Table 4.4.b

Here, the participant seemed to enjoy searching for books and bagged a total of nine books. In selection process of 1st book, browse function was used with keywords science, fiction and fantasy. Description of first book was undergone but the user choose second book without going through any metadata. For 2nd, 3rd, 5th and 6th book, same topics were searched and again without usage of metadata books were chosen. Again under same topics 4th book was selected but the user went through two books but the second one was chosen. In comment box, the user had mentioned that he/she wanted to read all books from 2nd to 6th in near future. As for 7th book selection, literature and fiction were browsed and furthermore scientific fiction section was selected. Also again user selected 8th and 9th book under same topic without usage of metadata as before whereas for 9th book, under browse function entertainment topic was selected and book was chosen.

4.5 Participant no.5

Task 1: Here interface 1 was used.

Sub-task	Books searched	Descriptions	Reviews	Publication	Tags	Book Bag
A	1				Y	2nd book selected
	2	Y			Y	
	3		Y			
B	1	Y			Y	Selected
C	1				Y	3rd book selected
	2					
	3					
D	1		Y		Y	Selected
E	1		Y			Selected

Table 4.5.a

Sub-task A: The participant used search functionality and went under non-fiction category to find books for surviving in island. User tags were seen in 1st book but user changed to another

book where descriptions were reviewed. But participant searched for 2nd book and reviewed descriptions and tags. Then again topic from user tags of 1st book was used to search 3rd book where the participant seemed to be interested in reviews. Though he/she might have thought 2nd book was more useful so finally selected the 2nd book.

Sub-task B: For this book, description and tag of book was reviewed and then simply selected.

Sub-task C: First the topic history was typed in and then user tag of 1st book was browsed. Among the listed books, 2nd book was then browed and selected but again cancelled. Finally user selected 3rd book without going into any details.

Sub-task D: Keyword knitting was punched in and a book was clicked. Reviews and user tag were respectively evaluated and selected.

Sub-task E: Book selection was done on the basis of author's name. Review was seen and the book was chosen.

Task 2: Here interface 1 was used.

No. of books	Books searched	Descriptions	Reviews	Publication	Tags	Book Bag
1	1	Y			Y	Selected
2	1	Y				Selected
3	1	Y				2nd book selected
	2	Y				
4	1	Y			Y	4th book selected
	2					

Table 4.5.b

For task 2, a total of four books were selected. In selection procedure of 1st book, literature and fiction category was viewed. Here user tags and descriptions were gone through respectively and the book had a rating of 5 stars so it was selected. 2nd and 3rd book were selected under same topic though in case of 3rd book, two books were reviewed but the second one was selected. The 4th book was chosen by using writing authors name directly to the search function. Two books were viewed by the user. Description and tags were seen but without using any functions the user selected second one, because it was a popular book and perhaps previously known by the user.

4.6 Participant no.6

Task 1: Here interface 1 was used.

Sub-task	Books searched	Descriptions	Reviews	Publication	Tags	Book Bag
A	1				Y	6th book selected but again cancelled as the book was selected from the data below.
	2				Y	
	3				Y	
	4	Y	Y		Y	
	5				Y	
	6		Y		Y	
	7	Y	Y	Y	Y	12th book selected
	8	Y			Y	
	9		Y			
	10		Y	Y	Y	
	11		Y			
	12		Y		Y	
B	1		Y			2nd book selected
	2		Y			
C	1				Y	Selected
D	1		Y			Selected
	2		Y			
E	1		Y			Selected

Table 4.6.a

Sub-task A: Survival in the desert was typed in search box and first three books with user tags were viewed. The user moved to 4th book where he/she viewed user tags, reviews and descriptions. Same user tag was used for 5th book also. Finally 6th book was selected after viewing user tags and reviews. Though the book selected in above process was deleted and inspection of another book was started. Descriptions, publication date, reviews and user tags were gone through respectively in case of 7th and 10th book whereas description and user tags

were seen in 8th book. In 9th book reviews were seen whereas in 11th book tag from 10th book was used. Finally 12th book was selected after going through reviews and user tags respectively.

Sub-task B: For selection of book regarding teaching something new, the participant used keywords like advance math for dummies and two books were seen along with their reviews. The 2nd book was chosen.

Sub-task C: In case of personal hobby and interest, keyword cat was punched in and a book was selected after going through user tags.

Sub-task D: User tag from sub-task C was used and keyword fiction was seen. Two books were evaluated by reading reviews and 2nd book was chosen.

Sub-task E: Again user tag from previous selected book was used within the same topic. Reviews were followed and book with 5 stars was selected.

Task 2: Here interface 1 was used.

No. of books	Books searched	Descriptions	Reviews	Publication	Tags	Book Bag
1	1				Y	3rd book selected
	2	Y	Y		Y	
	3	Y				
2	1	Y			Y	3rd book selected
	2	Y			Y	
	3	Y			Y	
3	1		Y		Y	Selected
4	1	Y			Y	Selected
5	1		Y		Y	Selected
6	1	Y			Y	3rd book selected
	2		Y		Y	
	3		Y		Y	
7	1	Y			Y	Selected
8	1				Y	6th book selected
	2		Y			
	3		Y			
	4				Y	

	5				Y	
	6				Y	

Table 4.6.b

In this non goal oriented task, a total of eight books were selected. In the category of literature and fiction, tag of first book was observed. For second book description, user tag and review were gone thoroughly. Then keyword feminism was typed in search functionality and description was viewed. Finally third book was selected. The 2nd book was selected by going through description and user tags for three books but the last one was chosen. In case of 3rd, 4th and 5th book, the participant seemed to love cats so the keyword cat was typed in. The book was selected after going through user tags and reviews. For the selection of 6th book, same category as above was used and three books were thoroughly reviewed. Descriptions and user tag was seen in first book whereas reviews and user tags were seen in second and third book. At the end third one was selected as 6th book. Again under same topic 7th book was chosen after evaluating description and user tag. The user tag of 7th book was used to explore 8th book. Six books were thoroughly reviewed. User tags were seen in first, fourth, fifth and sixth book. Reviews were viewed for second and third book. Finally the last book was chosen.

4.7 Participant no.7

Task 1: Here interface 2 was used.

Sub-task	Books searched	Descriptions	Reviews	Publication	Tags	Book Bag
A	1	Y				2nd book selected
	2	Y	Y	Y	Y	
B	1	Y				2nd book selected
	2	Y	Y	Y	Y	
C	1	Y	Y	Y		Selected
D	1	Y	Y	Y		Selected
E						Not Selected

Table 4.7.a

Sub-task A: Firstly keyword was searched in this task. After display of list of books, 1st book description was viewed. Then description, publication date and user tag were evaluated thoroughly. Even though the user did not find review of this book he/she selected this 2nd book based on other metadata.

Sub-task B: Two books were reviewed during selection procedure. Only description was followed for 1st book. Then user switched to another book under sci-fi category where he/she assessed descriptions, publication date, reviews and user tags.

Sub-task C: Under the category arts and photography, participant went through descriptions, publication date and reviews and thus selected the book.

Sub-task D: Search started under same category as above and among list of books user chose a book and went through description, publication date and reviews. After detailed analysis, user selected this book.

Sub-task E: In this sub-task, the participant did not choose any book even after browsing into search function with keyword fiction.

Task 2: Here interface 2 was used.

No. of books	Books searched	Descriptions	Reviews	Publication	Tags	Book Bag
1	1	Y	Y	Y	Y	2nd book selected
	2	Y	Y	Y		
2	1	Y				6th book selected
	2	Y	Y			
	3	Y				
	4	Y	Y		Y	
	5	Y				
	6	Y	Y	Y		
3	1	Y	Y			Selected

Table 4.7.b

In this task, three books were finalized by the user. For selection of 1st book, two books were evaluated at beginning. Descriptions, publication date and reviews of first book were seen. The user tried to see user tag as well but could not find. So he/she shifted to second book and viewed description, publication date and reviews and selected the book. In case of 2nd book, a total of

six books were evaluated by participant. Descriptions of all six books were viewed by the user. Review of second book was seen. Review and user tag of fourth book was viewed. Finally sixth book was selected on the basis of metadata evaluated. The 3rd book was selected under topic wheel of time on basis of description and review.

4.8 Participant no.8

Task 1: Here interface 1 was used.

Sub-task	Books searched	Descriptions	Reviews	Publication	Tags	Book Bag
A	1	Y				4th book selected
	2	Y	Y			
	3	Y	Y			
	4	Y	Y			
	5	Y			Y	
B	1	Y			Y	Selected
C	1		Y			2nd book selected
	2	Y	Y		Y	
	3	Y	Y			
	4	Y	Y			
D	1	Y	Y		Y	Selected
E	1		Y			7th book selected
	2	Y				
	3	Y				
	4	Y				
	5		Y		Y	
	6		Y			
	7		Y			

Table 4.8.a

Sub-task A: Combination of three surviving, desert and island were used in search box. In this task, total of six books were observed. For 1st book, description was viewed whereas for 2nd,

3rd and 4th books description and reviews were viewed respectively. Finally 5th book was selected after follow up on description and user tag.

Sub-task B: Keyword weaving basket was written in search box and description of book was viewed and then book was selected.

Sub-task C: For selection of book, review of 1st book was observed under category knitting. But user shifted to other books. Descriptions and reviews of 2nd, 3rd and 4th books were evaluated and selected.

Sub-task D: For selection of book in this sub-task, function filter by topic was used and world literature was typed in. Description was viewed which contains review from amazon.com and review was seen.

Sub-task E: Topic read was chosen from filter by user tags. Review from 1st book was observed. For 2nd, 3rd and 4th book only descriptions were viewed. The user moved to other books. Reviews of 5th, 6th and 7th books were seen. Finally the user selected 7th book.

Task 2: Here interface 1 was used.

No. of books	Books searched	Descriptions	Reviews	Publication	Tags	Book Bag
1	1	Y	Y			Selected
2	1		Y			7th book selected
	2	Y	Y			
	3		Y	Y		
	4		Y			
	5	Y	Y			
	6	Y				
	7	Y				
3	1	Y	Y			2nd book selected
	2	Y				
4	1		Y		Y	Selected

Table 4.8.b

Here four books were selected in the process. For selection of 1st book, description and reviews were seen. A total of seven books were evaluated in process of selection of 2nd book. Reviews of first, third and fourth books were observed. As for second and fifth book descriptions and

reviews were followed up whereas only description was noticed in sixth and seventh book. Finally seventh book was chosen. In case of 3rd book, two books were evaluated. Description of first and second book was seen and in addition review of first book was viewed. But the second book was chosen. For 4th book, review and user tag was followed up and chosen.

4.9 Participant no.9

Task 1: Here interface 1 was used.

Sub-task	Books searched	Descriptions	Reviews	Publication	Tags	Book Bag
A	1	Y				Selected
B	1	Y				4th book selected
	2	Y				
	3	Y			Y	
	4	Y				
C	1	Y				4th book selected
	2	Y				
	3	Y				
	4	Y				
D	1		Y		Y	6th book selected
	2		Y			
	3	Y				
	4	Y				
	5	Y	Y			
	6	Y	Y			
E	1	Y	Y			10th book selected
	2	Y				
	3			Y		
	4			Y		

	5		Y	Y		
	6			Y		
	7			Y		
	8		Y	Y		
	9	Y		Y		
	10		Y	Y		
	11		Y			
	12		Y			

Table 4.9.a

Sub-task A: As usual keywords like Desert Island and Surviving was written in search box and under the category biographies and memoirs chosen from filter by topic, desired book was selected on basis of its description.

Sub-task B: Here topic non-fiction was used from filter by user tag option. Among list of books four books were viewed. Description of 1st, 2nd and 4th book was viewed whereas description and review were viewed for 3rd book. Nevertheless 4th book was selected.

Sub-task C: Here category of science fiction and fantasy was selected where the user choose to observe four books again. Descriptions of all books were followed up but 4th book was selected.

Sub-task D: Here six books were thoroughly evaluated before the selection of 6th book. Keyword yoga was used for viewing books. For 1st and 2nd book reviews were seen whereas for 3rd and 4th book descriptions were viewed. In case of 5th and 6th book, description and reviews were followed.

Sub-task E: Different keywords were searched for selection of this book such as marine biology, massage, self-massage and health. Under keyword marine biology description and review were observed whereas under topic massage four books were viewed. : 2nd book's description, 3rd and 4th books publication date, 5th books publication date and reviews. Another four books were viewed under self-massage topic. Publication date of 6th, 7th, 8th, 9th, and 10th books were noticed. Also description of 9th book was viewed. Reviews of 8th, 11th and 12th books were noticed. Finally 12th book was selected.

Task 2: Here interface 1 was used.

No. of books	Books searched	Descriptions	Reviews	Publication	Tags	Book Bag
1	1		Y			Selected
2	1		Y			Selected
3	1		Y			Selected
4	1		Y			Selected
5	1		Y			Selected

Table 4.9.b

For task 2, five books were selected in total and all of them were selected on the basis of reviews. So, the user seemed to rely mainly on reviews for selection.

4.10 Participant no.10

Task 1: Here interface 2 was used.

Sub-task	Books searched	Descriptions	Reviews	Publication	Tags	Book Bag
A	1	Y				2nd book selected
	2	Y		Y		
B	1	Y				Selected
C	1					Selected
D	1	Y				3rd book selected
	2	Y				
	3	Y				
E	1		Y		Y	Selected

Table 4.10.a

Sub-task A: Keyword desert island survival guide was written in search box and browsed in this task. Descriptions for both books were seen and publication date for only 2nd book was viewed. The 2nd book was selected at the end.

Sub-task B: For selection of book, keyword basket played main role in search functionality where description was observed and thus selected.

Sub-task C: The book was selected after typing in gardening keyword in search box without going through any of the metadata.

Sub-task D: Three books were observed during selection process. The user went through description of all books under the same category as in sub-task C. Finally 3rd book was selected.

Sub-task E: Reviews and user-tags were seen for selection of this book under the keyword dogs.

Task 2: Here interface 2 was used.

No. of books	Books searched	Descriptions	Reviews	Publication	Tags	Book Bag
1	1	Y				4th book selected
	2		Y			
	3		Y			
	4		Y			
2	1	Y				Selected
3	1	Y	Y			2nd book selected
	2	Y	Y		Y	

Table 4.10.b

Three books were selected for task 2. For 1st book four books were went thoroughly in which description of first book and reviews of second, third and fourth under topic meals from the browse function was observed. After close analysis fourth book was chosen. As for 2nd book keyword low carbohydrate diet for cats was searched and after looking into one of the books description, it was selected. For 3rd book selection, description and reviews of two books were observed under category special diet. The second book was selected.

4.11 Participant no.11

Task 1: Here interface 1 was used.

Sub-task	Books searched	Descriptions	Reviews	Publication	Tags	Book Bag
A	1	Y				7th book selected
	2	Y				
	3	Y	Y			
	4		Y		Y	

	5	Y			Y	
	6	Y			Y	
	7	Y	Y			
B	1	Y			Y	3rd book selected
	2		Y			
	3	Y	Y	Y		
	4		Y			
C	1	Y	Y		Y	5th book selected
	2	Y				
	3	Y	Y			
	4	Y				
	5	Y	Y			
D	1				Y	2nd book selected
	2		Y		Y	
E	1	Y	Y	Y	Y	Selected

Table 4.11.a

Sub-task A: Search keywords wilderness survival and desert survival were used in order to select the book. Seven books were browsed where descriptions for 1st and 2nd book used. Descriptions and reviews for 3rd and 7th book were seen and only reviews for 4th book is used. For 5th, 6th and 7th book descriptions and tags were seen and 7th book was chosen.

Sub-task B: Chinese business culture was keyword used to search the book and four books were seen thoroughly. Only description was seen for 1st book and only reviews for 2nd book. Descriptions, reviews and publication date were observed for 3rd book whereas reviews were viewed for 4th book. Finally, 4th book was selected.

Sub-task C: Long range riffle shooting was searched for this book and five books were observed. Descriptions, review and tags were respectively observed for 1st book. Only descriptions were used for 2nd and 4th book whereas descriptions and reviews for 3rd and 5th book. Finally 5th book was chosen.

Sub-task D: User tags topic was used from the book selected in sub-task C and 1st book was seen where topic from user-tags was clicked. Among the displayed books, 2nd book was selected based on reviews and user-tags.

Sub-task E: Here search keyword death was used to find the book. Among listed books user chose a single book where he/she observed descriptions, publication date, reviews and user-tags respectively. No reviews were found but book was selected on the basis of other metadata.

Task 2: Here interface 1 was used.

No. of books	Books searched	Descriptions	Reviews	Publication	Tags	Book Bag
1	1	Y	Y		Y	Selected
2	1	Y	Y		Y	Selected
3	1	Y	Y	Y	Y	Selected
4	1	Y	Y			Selected
	2	Y	Y			

Table 4.11.b

Four of the books were selected for this task 2. For 1st book skill development was the keyword used for search. Descriptions, reviews and user-tags were seen and the book was selected. Under same keyword other books were also viewed. Descriptions, reviews and user-tags were observed and thus 2nd book was selected. 3rd book was selected by viewing its descriptions, publication date, reviews and user-tags respectively. For 4th book search keyword negotiation was used where descriptions and reviews were seen for first and second book but finally second book was selected.

4.12 Ratings given by all the participants for task 1 and task 2:

Ratings were given according to the usage of descriptions, reviews, user-tags and publication date by eleven participants from I to V. These ratings are illustrated below in the tables:

Ratings of task 1

Ratings	I	II	III	IV	V
Descriptions		1	2	5	3
Reviews			1	4	4
User-tags	1		1	5	
Publication date	1	1	3	1	

Table 4.12.a Ratings given for task 1

Ratings of task 2

Ratings	I	II	III	IV	V
Descriptions		2		6	3
Reviews		1	2	4	4
User-tags	1	2	3	2	
Publication date	4	1	2	1	

Table 4.12.b Ratings given for task 2

I – Barely used II – Very rarely used

III – Rarely used IV – Often used

V – Very Extremely used

In the tables maximum number of ratings V is given to reviews and descriptions. Reviews are rated by more number of participants than descriptions in this section which are respectively 4 and 3 as seen in the tables 4.12.a and 4.12.b. In case of task 1, rating IV, the most can be viewed in descriptions and user-tags done by 5 participants whereas in review done by 4 participants. Only 1 participant had chosen publication date. As for task 2 descriptions is rated IV by 6 participants, reviews by 4 participants, user-tags by 2 participants and publication date by only 1 participant as in table 4.12.b. Ratings like III, II and I for all kinds metadata were given by very few participants as shown in both tables above. Thus description and reviews appear to be ultimate choice of users. User-tags are comparatively preferred by participants more in task 1 rather than task 2 whereas publication date is least preferred in both tasks.

Chapter 5: Findings and Discussion

From these user's data for a book search, which was based on two interfaces and two tasks, usage of professional metadata and user-generated content depended on the task given.

The sub-task A that comes under task 1 had the requirement to select a particular book which helps survival in the desert. Here considerations of various options while data analysis is discussed. Every user's utilized search functionality to write a query related to book search regardless of the interface. The tables depicted a sequence where users observed description firstly then review secondly whereas user-tags and publication date were rarely used. Thus, based on these analysis commonly used metadata were description and reviews for search of particular book.

In sub-task B the user had to select book that teaches them something new. So, on this task once more search functionality was used often where the queries were related to this task. Descriptions and reviews were mostly viewed to finally select the book that suits them to teach something new. Also, in some case users used browsing function under various topics that falls under filter by topic and filter by user tags option in both of the interfaces though the difference between these two interfaces were that in one interface it contained lots of books in browse function while in other interface this function was not included. Though at times only search functionality was used and selection of books did not depend on metadata which can be seen in Table 4.1.a for the sub-task B. Nevertheless user-tags were also utilized more in this case than the previous sub-task A which can be seen in the tables 4.4.a, 4.5.a, 4.7.a and 4.8.a. The publication dates were used by very few users as seen in all the tables above in data analysis section of task 1 tables. Also what can be seen in the comment box by most users were that description and reviews helped them the most to select this book which can be seen in the figure 5.b as shown below.

The third sub-task C under this category was to find book about personal hobbies or interest. So for this case also the most used metadata were the description and reviews but what else was also used by some users were the publication dates as well which can be seen in tables 4.3.a, 4.4.a and 4.7.a.Very few users used the user-tags which is illustrated by the task 1 tables above in data analysis part. In comment box most of the users had written that reviews and descriptions assisted them to find desired book.

The sub-task D in which the user have to select book based on the previous book user's recommendation and ratings. So for this task obviously most of the user used reviews as seen in almost all the tables. But some also used descriptions and user-tags whereas only few used publication date. In analysis period, I found that some users had highlighted about user ratings in the comment box as depicted in figure 4.3.b, participant no.3 table for task 1. In addition other user had commented about reviews from amazon.com which was helpful during selection of book as seen in figure 4.2.a of participant no.2 in the data analysis section.

The fifth sub-task which guided users to find a particular book for fun and during analysis this task I found that yet again users had generally utilized description and reviews most of the time and very few had utilized user-tags and publication date. During this task two interesting comments regarding cover pages of books were posted by users which can be seen in figure 5.a below.

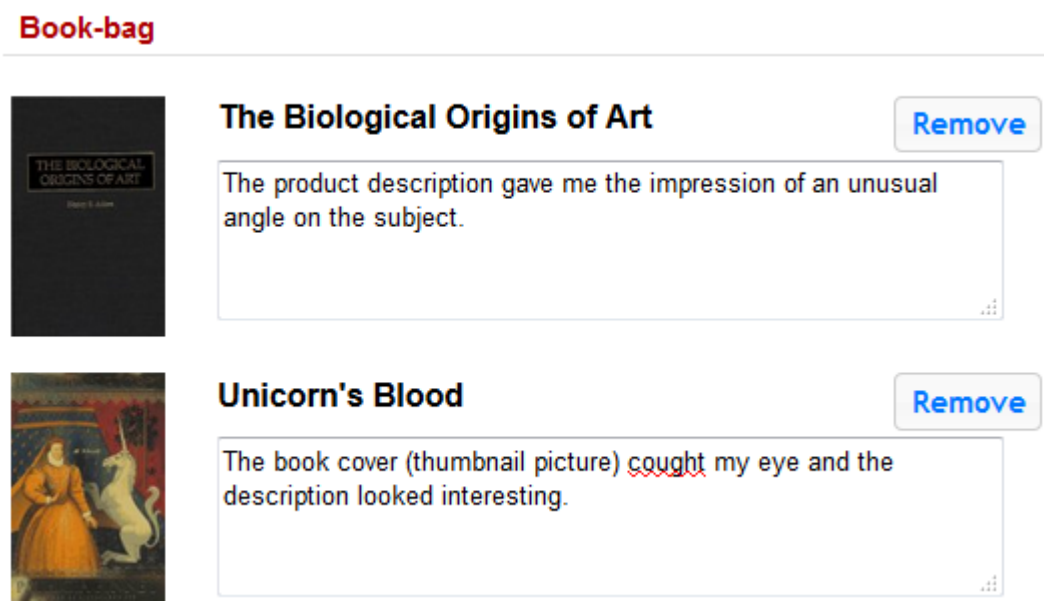


Figure 5.a. Snapshot of comments written after selection of the book.

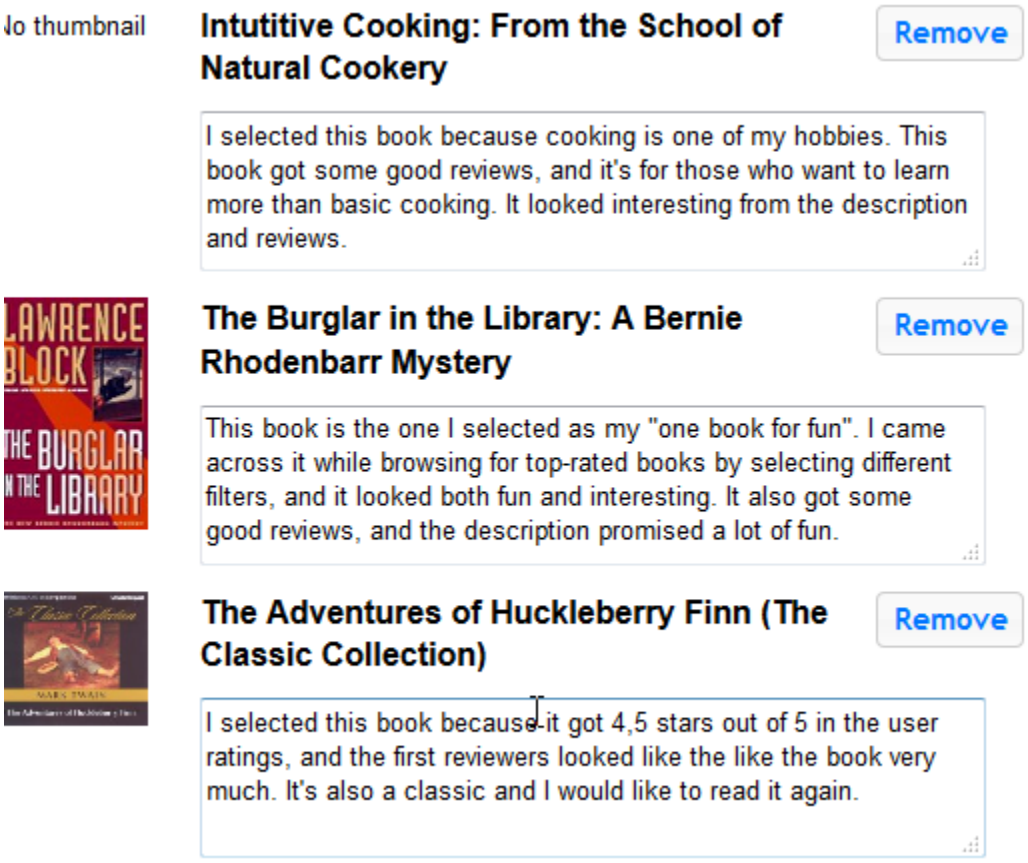


Figure 5.b. Snapshots of comments from the participant for some books.

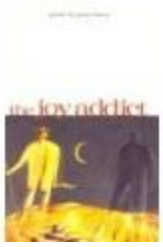
In task 2 where users can browse website as long as they are not bored and search for interesting books while waiting for a friend. Here the users selected up to maximum of eleven books which can be seen on data analysis section by participant no. 3 and minimum of three books. Most of them selected up to four books and few selected six or seven books which can be seen in the tables of task 2 from data analysis section. The users managed to select these books on the basis of descriptions and reviews including user-tags. Only few of them chose publication dates although some users have mentioned that publication date is equally important to know latest edition of book as shown in figure 5.c. Patterns were seen in some users while selection of particular books where all started from description and then reviews. An example of it can be seen in table 11.2.b in task 2 done by participant 11.

Comments on the interface 1 were also seen from one of the participant which stated it had limited features for popular books and there was no recommendation features which is shown in the figure 5.d.

Emotional Intelligence: Achieving Academic and Career Success Remove

What time the book was published matters: evaluating its usefulness will for some subjects (which changes With fashion) be dependent on this. Also, the amount of reviews matters, but only if the book isn't new. I'm left to interpret the actuality of the book by it's literal cover.

Figure 5.c. Snapshots of comment by the participant for the importance of publication date.



The Joy Addict (Carnegie Mellon Poetry) Remove

As I can't search for most popular books in this Interface, and there is no recommendation feature, I opted for finding anything that was tagged as "wonderful" (suggesting enthusiasm) and had good reviews (measured in amount of stars...)

Figure 5.d. Snapshots of comment by the participant for interface 1.

Based on the literature which states metadata helps users to find relevant data as quick as possible, so based on this review I found similarities on the retrieval of book.

Folksonomy is described in literature which declares user-tags generated by users and helps in finding necessary and related information. So here folksonomies which are generally user-tags helps users to search for related topics and hence assist in finding particular book.

For book to be searched, the search functionality is used in which user enters related keywords and then starts searching but in some case when user enters keywords and indexed words do not match then there are some issues of not finding correct information as expected. Thus in some cases while analyzing these data I found some similar problem of not finding required information rather it says no match found, so what can be observed from this is that user has to enter correct type of keyword or combination of keywords. Sometime it's quite confusing to find particular words that are described in index making it difficult to find information.

Chapter 6: Limitations and future work

Due to limited source of data collection, it was complex to generalize the problem while retrieving essential information and usage of metadata could not be interpreted more effectively.

Another limitation is the software Morae Manager mode in which data were analyzed. As this software was a trial version and could only be used for 30 days so restriction in time also created a drawback to go through user's data in detail.

Furthermore, the databases were only taken from Amazon and library thing. The data were limited to these databases only. If there were several other databases then the user might get some more comfort in retrieving necessary information.

Future work could be implemented by addition of social book search data in a transaction log analysis for better analysis and that helps in generalizing the research as a whole.

Chapter 7: Conclusion

In RQ.1 all the participants utilized both of professional metadata and user-generated content, especially description and reviews were mostly used by all in both of the tasks and user-generated tags were applied by some users whereas publication date was rarely utilized compared to the previous one. From this I could analyze that for any book search the need of effective and efficient professional metadata and user-generated content plays a crucial role in searching relevant information within short period of time. Nevertheless prior knowledge of books and the literature field has resulted in less utilization of professional metadata and user-generated content.

For RQ.2 the participants were given mandatory goal oriented task 1 to select book and their behavior were analyzed. In sub-task A search functionality was used the most in both interfaces. For sub-task B search functionality was majorly utilized in interface 1 while topic under browse function as well as search functionality was used in interface 2. In sub-task C participants used search keywords and explore their favorite topics in both of interfaces. In sub-task D participants were busy reading reviews and ratings as narrated by the task to select book based on recommendation. In sub-task E, various topics under both of interfaces were used as well as search functionality was used to select the book.

For non-goal oriented task 2 all participants were searching on their topic of interest. Some participants chose as many books as they could have and others chose a few depending on interest in exploring the system with 2 interfaces. Hence, in this task it was visible that some participants seemed to enjoy search and selection book.

The behavior of participants is affected by types of task, as seen in these two tasks. Goal oriented task was enjoyed least compared to non-goal oriented task. As most of participant's selected more than 3 books in task 2 so it can be said that most of them seemed to enjoy task 2 rather than task 1.

In case of RQ.3 interface 1 has limited content like filter by topic and filter by user-tags under which inadequate content topics were listed. So mainly search functionality were used to search particular book and very few utilized filter by topic and filter by user-tags. But in interface 2 apart from features included in interface 1 there is one extra feature which is browse function that was used by participants. It seemed that additional features in interface 2 were fully utilized by participants while exploring books on diverse topics.

References

Bogers, T. & Petras, V. (2015), 'Tagging vs. controlled vocabulary: Which is more helpful for book search? ', *In iConference 2015 Proceedings*.

Conradi, E. (2009), to_be_classified: A facet analysis of a folksonomy, Master's thesis, Høgskolen i Oslo.

Gäde, M., Hall, M. M., Huurdemann, H., Jaap, K., Koolen, M., Skov, M., Toms, E. & Walsh, D. (2015), 'Overview of the SBS 2015 interactive track', *CEUR Workshop Proceedings* **1391**.

Hammond, T., Hannay, R., Lund, B. & Scott, J. (2005). 'Social bookmarking tools (I)', *Nature Publishing Group* **11**(4).

Koolen, M., Bogers, T., Kazai, G., Kamps, J. & Preminger, N. (2014), 'Overview of the INEX 2014 social book search track', http://ceur-ws.org/Vol-1180/CLEF2014wn-Inex-KoolenEt2014.pdf. [Accessed: Aug. 2015]

NISO (2001),'Understandingmetadata',http://www.niso.org/publications/press/UnderstandingMetadata.pdf. [Accessed: Jun. 2015]

Nordlie, R. & Pharo, N. (2012), Seven years of INEX interactive retrieval experiments – lessons and challenges, in T. Catarci, P. Forner, D. Hiemstra, A. Peas & G. Santucci, eds,' Information Access Evaluation. Multilinguality Multimodality, and Visual Analytics', Vol. 7488, Springer Berlin Heidelberg, pp 13-23.

Nordlie, R. & Pharo, N. (2013), Search transition as a measure of effort in information retrieval interaction, in 'Proceedings of 76th ASIS&T Annual Meeting: Beyond the Cloud: Rethinking Information Boundaries', ASIST '13, pp. 63:1-63:7.

Peters, I. (2009). *Folksonomies Indexing and Retrieval in Web 20*. (1st Ed.). Germany: Deutsche Nationalbibliothek.

Pharo, N., Nordlie, R., Fuhr, N., Beckers, T. & Fachry, K.N. (2010), 'Overview of the INEX 2009 Interactive Track', https://oda.hio.no/jspui/bitstream/10642/372/2/531882.pdf. [Accessed: Oct. 2015].

Strasser,C. (2015). 'ResearchDataManagement'. http://www.niso.org/apps/group_public/download.php/15375/PrimerRDM-2015-0727.pdf. [Accessed: Sept. 2015].

Trant, J. (2008). 'Studying Social Tagging and Folksonomy: A Review and Framework', *Journal of Digital Information*.