Le Thi Thuy Phuong

Processes of tagging behaviors: Case studies of CiteULike, LibraryThing, YouTube and Fickr

Supervisor: Prof. Nils Pharo

Oslo University College

Faculty of Journalism, Library and Information Science

DECLARATION

I certify that all material in this dissertation which is not my own work has been identified and that no material is included for which a degree has previously been coffered upon me.

……Le Thi Thuy Phuong…… (Signature of candidate)

Submitted electronically and unsigned

ACKNOWLEDGEMENTS

This work is finished with the supports from many persons in DILL program and some other friends.

I am very grateful for the critical advice and supports from my supervisor, Professor Nils Pharo. He has take time to follow each researching step to give right comments and helped me to go in right ways. Besides, He was very patient to read and comment on each idea, grammar to make my work become better.

I would like to take this chance to thank my informants in participating the long experiments and interviews. Without the enthusiasm, I cannot have enough data to finish my thesis.

Specially, I would like to say thankfulness to Professor Ragnar Nordlie because he gave me many useful comments in the first seminar and second seminar. According to those comments, I changed some ideas in data collection and data analysis.

Last but not least, I am happy with my friends, Chau and Vi who commented, gave ideas and read, edited grammars on my work.

ABSTRACT

This work is a study of tagging behavior processes in different types of documents and tagging sites. Moreover, the tagging challenges are addressed following tagging behavior processes. The study is limited to CiteULike, LibraryThing, YouTube and Flickr as case studies. However, tagging behavior processes and tagging challenges are concentrated rather than evaluation of the above tagging sites.

To define the processes of tagging behavior and the tagging challenges, the talk aloud experiments have been conducted to the informants who are real taggers in various tagging sites such as Facebook, Flickr, YouTube, Picasa as well as some other personal blog. The surveys of tagging experience and open interviews have also taken a long with the experiments.

The results of this study show that there are 2 main types of taggers: confident taggers and certainty taggers. Those types of taggers have different behaviors and feeling during tagging. Moreover, the process of video and image tagging is distinguished with text tagging. Most of the informants felt simpler and easier to tag for pictures and videos than text since the videos and pictures are visible and easier to understand. According to the processes, there will be 4 main tagging issues such as information resources issues, tagging function issues, vocabularies issues and patience in tagging.

The findings in the thesis could be a useful reference for building tagging systems in practices. Besides, the thesis might be used as hypothesis of tagging behavior to other relevant researches.

Keywords: tagging behavior processes, user indexing, tagging challenges, tagging factor influences.

TABLE OF CONTENT

ACKNOWLEDGEMENTS 3
ABSTRACT 4
CHAPTER 1: INTRODUCTION 9
1.1 Background 9
1.2 Research problem statement 10
1.3 Research questions 10
1.4 Research objectives: 10
1.5 Research methodology 11
***1.6* Scope of the research** 11
1.7 Thesis outline 12
CHAPTER 2: PROCESS OF TAGGING BEHAVIOR: STATE OF ART 13
2.1 Define of tagging 13
2.2 Value of tag 14
2.3 Process of cognitive tagging 15
2.4 Information resources in tagging 17
2.5 Motivation of tagging 18
2.6 Cultures of tagging 20
2.7 Types of tags 20
2.8 Semantics of tag 21
2.9 Solutions for tagging issues 22
CHAPTER 3: RESEARCH METHODOLOGY 24
3.1 Research methodologies 24
3.2 CiteULike, LibraryThing and Flickr, YouTube as case studies 25
3.3 Sampling technique 26
3.3.1 Size of samples 26
3.3.2 Criteria for recruitment 26
3.4 Data collection 26
3.4.1 Objectives 26
3.4.2 Tasks designs 27
3.4.3 Instructions 28

3.4.4 Instruments 29
3.4.5 Moderator 29
3.4.6 Process of experiment 30
3.4.7 Observation design 30
3.4.8 Interview design 31
3.5 Pilot study 31
3.6 Data analysis methods 31
3.7 Ethical consideration 32
CHAPTER 4: DATA ANALYSIS 33
4.1 Geographic of informants 33
4.1.1 Tagging perception and tagging experience 33
4.1.2 Classification of taggers 35
4. 2 Explorations 36
4.2.1 Exploration of confident taggers 37
4.2.2 Exploration of certainty taggers 38
4.2.2.1 Explorations within the tagging pages 38
4.2.2.2 Explorations outside of the tagging pages 41
4.2.2.3 Topic memorization 42
4.2.3 Explorations of video and image 42
4.2.3.1 Video and image memorization 43
4.2.3.2 Doubts about video and image 43
4.2.3.3 Summaries of exploration stage 43
4.3. Tagging decision 43
4.3.1 Tagging prediction through the match between tags and searching phrases 45
4.3.2 Tagging motivations 47
4.3.3 Tag formulation 48
4.3.3.1 Tag formulation of confident taggers 48
4.3.3.2 Tag formulation of certainty taggers 49
4.3.3.3 Tag formulation of video and image 50
4.3.4 Tagging consideration 51
4.4. Type of tags 52
4.5 Factors influence on tagging behaviors 57

4.5.1 Personal tendency 57
4.5.2 Communities' tags 57
4.5.3 Indexing skills 58
4.5.4 Tagging experience 59
4.5.5 Tagging objects 59
4.5.6 Cultural knowledge 60
4.5.7 Exploration of detailed metadata 60
4.6. Tagging challenges 61
4.6.1 Information resources issues 61
4.6.1.1 Unattractive information structure 61
4.6.1.2 Poor metadata 62
4.6.2 Tagging function issues 62
4.6.2.1 Quick tagging functions 62
4.6.2.2 Tag strings combinations 63
4.6.2.3 Poor suggested tags 63
4.6.3 Vocabulary issues 64
4.6.3.1 Misspellings 64
4.6.3.2 Difficulties in term understand 64
4.6.3.3 Difficulties in tags extractions 65
4.6.3.4 Ambiguous tags 65
4.6.*4 Patience issues* 67
CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS 70
5.1 Discussions of the research questions 70
5.1.1 Discussion of research question 1 70
5.1.1.1 Tagging behavior process of the confident taggers 70
5.1.1.2 Tagging behavior process of the certainty taggers 72
5.1.1.3 Tagging behavior process of video and image 75
5.1.1.4 Compare of the text and video, image tagging processes 76
5.1.2 Discussion of research question 2 76
5.2 Recommendations 77
5.2.1 Improving information resources 77
5.2.2 Improving tagging functions 77

5.2.3 Improving tagging support 78
5.2.4 Improving tag ambiguities 79
5.3. Research limitations 79
5.4 Further studies 80
REFERENCES 81
APPENDIX 1: TAGGING TASKS DESIGN 85
APPENDIX 2: GUIDELINE OF TALK ALOUD EXPERIMENT 86
APPENDIX 3: PREPARTIONS FOR THE EXPERIMENT 88
APPENDIX 4: QUESTIONNAIRES FOR THE PRE-EXPERIMENT 89
APPENDIX 5: INTERVIEW QUESTIONS 90
APPENDIX 6: TAG RESULTS 91
APPENDIX 7: TOPIC EXPERIENCE 94

CHAPTER 1: INTRODUCTION

The chapter presents background and statement of research problem as well as the aim of study and research questions. In addition, the scopes of study and research methods are presented. Finally, an outline of thesis is introduced.

1.1 Background

There are three types of metadata creators: professionals such as cataloguers, authors, and users (Mathes, 2004). Tagging is an approach of user created metadata. As the result, tagging has been implemented in many systems such as photos, videos, web pages, Wikipedia articles and academic paper citations. Since it helps user to navigate, to find, to refine and to share documents within communities (Chi, E. H. and Mytkowicz, T., 2007).

Tagging behavior is concerned with two main kinds of studies: tagging motivation, tagging cultures, linguistic of tag, semantic of tag; and topic memories, topic experience, tagging feeling, tagging consideration, tagging decision making and tag formulations.

The first kind of studies is tagging motivation, tagging cultures, linguistic of tag as well as semantic of tag. Dong and Fu (2010) found that for the tags of Europeans and Americans are more focused on main objects than the tags of Chinese. Ames and Naaman (2007) figured out 4 types of tagging motivations: self archives, self communication, social archives and social communication. For tagging challenges, Sinclair and Cardew-Hall (2008) found that tags are spreader out by synonyms, polysemy, and variants of tags, misspellings, abbreviations and slang of tags.

The second kind of studies is cognitive tagging process such as tagging perception, thought, feeling and decision making which have few papers study on it. Rashmisinha (2005) had mentioned cognitive tagging process in an academic blog post which is mentioned in some academic papers and a master research of Marvasti (2008). The tagging process is discussed under 2 cognitive tagging stages: personal tendency and tagging decision. However, because of short conversation and debate, it is deserved to study both cognitive tagging process and tagging behavior process.

1.2 Research problem statement

Wu, Chao, Zhou and Bo (2009) have defined three major dimensions in tagging: user to resources (resources tagged by users); tag to resources (many sources connected); user to users (share tags, tag suggestions). Recently, most of the studies are focused on the approach: tags to resources, such as types of tags, semantic of tags and tagging applications to systems. In the contrast, this study will be approached by user to resources dimension which means to user behavior to tag resources.

According to the research background, the tagging behavior and cognitive tagging are necessary to be investigated and discussed together in a study, since they have relationships. Firstly, many studies in tagging behaviors in modern days concentrate on tagging motivations, value of tags, structures of tags, types of tags, cultures of tags as well as some applied researches about compare of folksonomies and taxonomies, tagging application in recommendation systems, etc. Those studies are investigated separately while most of them have relationship in a process. For example, in different cultures may lead to various tagging motivations and type of tags or tagging behaviors. Secondly, the concept of tagging behavior could be understood not only as motivations or types of tags, but it also as thinking, feeling, and activities through process of tagging.

1.3 Research questions

Research question 1*: What are user behaviors characteristics in tagging processes?*

Research question 2: *Which challenges do the taggers meet during tagging processes?*

1.4 Research objectives:

- Investigating the different processes among different taggers and different types of objects such as text documents, videos or images.
- Understanding actions and behaviors in each step of the process.
- Exploring perception, thought, feeling, uncertainty and tag formulations from taggers.
- Figuring out tagging issues during process of tagging which to support taggers.

1.5 Research methodology

Since the topic relates to tagging behaviors and tagging cognitions, this study is based on qualitative to collect and analyze data. In order to collect and observe tagging behaviors, there are some possible approaches such as using real observations from the realities, interviews as well as experiments. According to many researches about cognitive psychologies, almost the studies are used think aloud experiments (talk aloud or talk-louder) to investigate user's cognitive, feeling as well as process of decision making (Van Someren, M. W., Barnard, Y. F., and Sandberg, J. A. C., 1994). In this study, talk aloud experiments will be the main method to collect data to answer for research questions. The talk aloud experiment mean to during the experimental tasks testing, informants try to say anything that goes through their minds. Besides, the study also uses observation, questionnaire survey before and after test and interview.

1.6 Scope of the research

The study will be approached by user to resources dimension which aim to observe how taggers behave through tagging with different objects. Furthermore, the observation of the study is going to start from topic experience, tagging experience, searching, and selection, motivation, tagging decision making and tag formulations in the tagging process. Refine of tags, search by tags and edit tags as well as other tagging activities are not considered in the study.

There two main concentrations of the study. Firstly, it is important to investigate and explore activities and behaviors through the tagging process. For example, in the exploration stage, the study will observe what tagger searches documents, what kind of things to be read by taggers, how they read and understand to tag as well as in the tag formulation tagger will be observed what, how and why they tag for documents. Secondly, the perceptions, thought, belief, knowledge, experience and feeling through the tagging process also is observed and analyzed. However, for the second purpose, there will be a thorny challenge that the cognitions are very difficult to observe and understand deeply. As the result, in a limitation, the study tries to investigate in by the experiment, talk aloud and interview questions.

1.7 Thesis outline

The content of thesis is presented in five chapters, thus to to table of content, figures and tables, reference and appendixes.

Chapter 1 presents background and statement of research problem as well as the aim of study and research question, brief introduction of research methodology and scope of study.

Chapter 2 delivers an important review of recent studies about various tagging behaviors, tagging cognitions as well as tagging challenges and solutions.

Chapter 3 provides the explanation of methods used in the research and the details of implementations in data collection and data analysis.

Chapter 4 discovers the data collected by analysis and discussions. Afterwards, findings of the research are summarized.

Chapter 5 approaches the conclusion and recommendation for the research. It provides answers for research questions set up at the beginning. Also, the chapter helps to solve the research issues and to conduct further research related to topics

CHAPTER 2: PROCESS OF TAGGING BEHAVIOR: STATE OF ART

This chapter presents studies of tagging behavior process from 2004 to now. Tagging behaviors such as type of tags, structures of tags, cultures of tagging, motivation a long with tagging cognition is discussed. The review is searched through ISIS-Web of Science, library database such as ACM Library, Springerlink, and Emeral Insight as well as through the Google with non academic resources such as academic blog posts. Furthermore, many keywords are addressed for data collecting such as: *tagging cognition, cognitions theories, user behavior of tagging, tagging behavior, personal ontology, distributed classification; ethno-classification, social annotation, user indexing and user-classification are used in searching*.

2.1 Define of tagging

Tagging sites have become more popular since 2004 (Hammond, Hannay, Lund, and Scott, 2005). There are three types of metadata creators: professionals (e.g. cataloguers), authors, and users (e.g. taggers). Tagging is an approach of user creating (Mathes, 2004). As the result, tag or tagging has been defined by many perspectives:

Firstly, tag is understood as free keywords from users. "*Tags are user-defined descriptors that can be any string assigned to target resources. They're a little bit like keywords but nonhierarchical, and can be freely chosen by the user without any apriori dictionary, taxonomy, or ontology to conform to.*" (Wu, Chao and Zhou, Bo, 2009; p.21)

Secondly, it is approached by relationship between three major tagging object such as tag, tagging object (book, article, video or image) and tagger. "*Tag is users' description on resources. It acts as the connection between people's subjective cognition and objective information*." Simultaneously, there are such kinds of connections: User to resources (resources tagged by users); Tag to resources (many sources connected); User to users (share tags, tag suggestions). (Wu, Chao and Zhou, Bo, 2009; p.21)

Thirdly, tag is mentioned as personal tendency such as knowledge, experience or belief about topic. *"Tag is an attractive feature of folksonomies is their inclusiveness; they reflect*

the vocabulary of the users, regardless of viewpoint, background, bias, and so forth." (Spitieri, L., 2007, p.460)

Sen, Lam, Rashid, Cosley, Frankowski, Osterhouse, Harper, and Riedl (2006) divided tag systems in two dimensions:

- Broader system which allows any users can do any tags for any items. LibraryThing is a case in point.
- Narrow system which just allows for content creators submitting their tags. For example of such systems are Flickr or Technorati.

According to these approaches, the process of tagging behavior from two kinds of taggers might be different by different tagging systems. Compared to taggers of broader tagging systems, taggers of narrow system may be sure of tags because they are tag's creator of their own videos and images.

2.2 Value of tag

Tagging has been implemented in many systems such as photos, videos, web pages, Wikipedia articles and academic paper citations as it helps user navigate, find, refine their documents and sharing related documents with communities (Chi, E. H. and Mytkowicz, T., 2007). On the other hands, Sinclair, & Cardew-Hall (2008) had discovered in an experimental study, which participants would evaluate between tags and search box that users preferred to search by search box for specific task than tags and tags were chosen for general search tasks. As the result, tags were useful for browsing (or non-specific information discovery), and for providing visual summary of the database. From tags, users could scan and choose suitable tags for searching both easily and quickly.

In conclusion, tags is now useful for recommendations or for sharing documents among users, but it also helps people to browse and to retrieve documents in system. However, tags will be not useful in some difficult searching tasks or in specific searching tasks. In regards of tagging behavior process, it is necessary to have two kinds of tasks in this study: the general search tasks and the specific search tasks, as each of them may drive to different process of tagging.

2.3 Process of cognitive tagging

Tagging behavior process is close to cognitive psychology process. The first is the perception which includes psychology, attention, time sensation as well as understanding. The second are memories such as tacit knowledge, explicit knowledge or experiences on topics, logic formulation, choice, uncertainty relate to tacit knowledge, judgment and decision making. The third are languages to show their tacit knowledge. People may have challenges to choose good tags because they want to share well-structure and correct grammar with other people. (Wales, J., 2010)

Rashmisinha (2005) has mentioned cognitive process of tagging in an academic blog post which is reviewed in a master research of Marvasti (2008) and received comments from many researchers and practitioners communities. The author has observed from realities, used her experience of tagging from users and cognitive psychology theories to show that:

There are two stages of cognitive tagging. The first stage is "*Related category activation*" which means human brain usually imagines to other things which related to the topics before tagging. This stage relates to perceptions and memories. For instance, to read Harry potter part 1, a tagger could think to topics, opinions about that book and motivation of tags, etc. The second step is *decision making*. In this step, tagger might make an adjustment between topic, experience and documents.

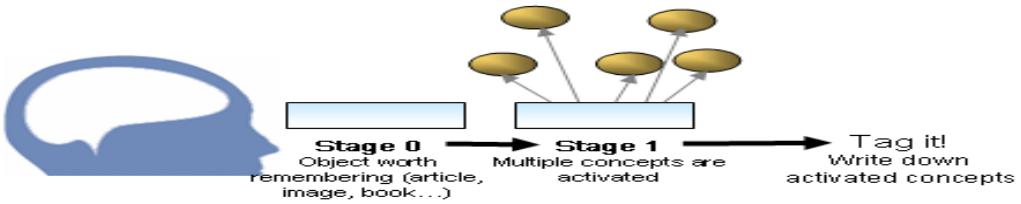

Figure 2.1: Cognitive tagging process (Rashmisinha, 2005, p. 1)

In practice, the process of cognitive tagging is very complicated. Therfore, the author argured with herself that in realities of digital environments, making decision is not easy because of consideration, choices and uncertainty. However, she didn't point out what kind of considerations such as tag corrections, number of tags or tag synoyms. She called this

complicated process is "*Post-activation analysis paralysis*". The step may be influenced by other factors such as: individual knowledge, cultures and the new knowledge of documents. In this stage, these factors could conflict each others. Consequently, they can be negotiated to give final decision for many level of goals (such as navigation, refindablitie). She said in this stage, taggers may feel fear to give wrong decisions which may take their time and enough patient to repair.

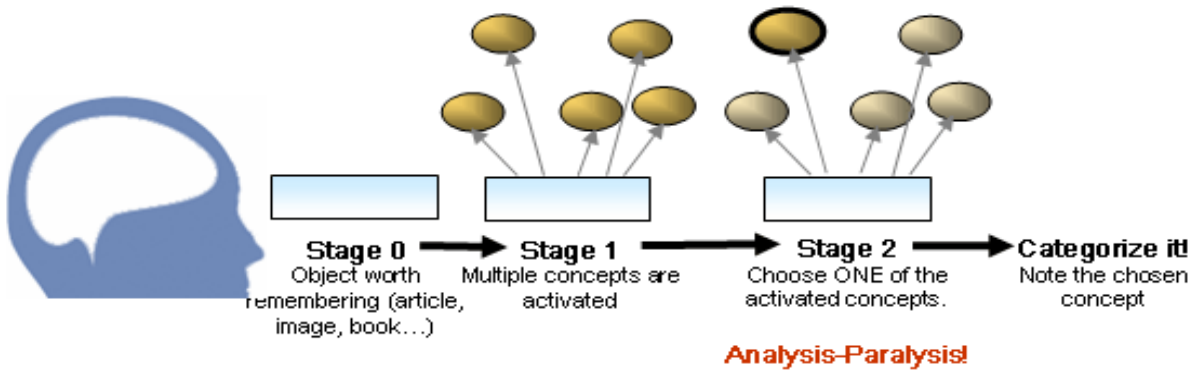

Figure 2.2: Whole stage of cognitive tagging process (Rashmisinha, 2005, p. 1)

After the posting, the essay has been received many comments from the tagging communities. Almost people agreed with her that the process was very interesting. It could become a good reference for the tagging behavior process. Nevertheless, in many comments, some enthusiasm commentators raised different interesting and skeptics perspectives such as:

- Commentators said Rashmisinha did not mention "input" of tagging process such as tagging objects (text or video or image). They argued that tagger may be lazy to spend time for thinking of tags so that it may drive to different processes. Not only have such kinds of tagger, but also for different cultures, languages and especially for different purposes of tagging drive to different process of tagging.
- Other commentators contributed that the selfishness impacts to tag results as taggers may want the tags reach to other taggers. They raised a question that "*Is*

there privacy issue in tagging, as it could have many sensitive tags which taggers do not want to share?"

Some important points in reading both essay and comments are addressable as follows:

Firstly, the cognitive tagging process still lacks of mentioned issues such as input and output of tagging process; internal and external factors influence on cognitive tagging process and different cultures of tagging make different tags results. The author simply mentioned the general cognitive tagging process without a concrete context such as tagging with fictions, factual topics and entertainment topics.

Secondly, two stages of cognitive process strongly focused on feeling and thinking more than tagging behaviors. Moreover, she still didn't discuss of tag formulations in the second stage. Consequently, the author mentioned to the fear of taggers to put tags in the "*Post-activation analysis paralysis"* stage but didn't address in which context tagger would be afraid of putting tags and how they overcome their fears to tag as a sequence.

Thirdly, it seemed the process only focused on internal process and it did not indicate external factors. For example, in the stage of topic memorizations, there were not only the object's reminder for related things but also taggers might be influenced by the tagging system as well as level of interest on objects.

To conclude, the process has been discussed in general. The author did not distinguish different processes in different contexts, different cultures and different tagging objects (print, video or images). For example, tagging process on tagger in narrow tagging system might differ to broad tagging system.

2.4 Information resources in tagging

Bar-Ilan, Zhitomirsky-Geffet, Miller and Shoham (2010) compared between three groups of image tagging to study the effectiveness of information background on tagging behavior. The first group saw only images without any information. The second group saw images with short descriptions as well as titles, etc. The final group showed short descriptions, and links to web pages which the image appeared. They found that the more read information

and interacted with the web page, taggers had more popular tags and number of tags. It means there is a relationship between tags and information background. However, the study was approached by examinations between tags and information background without consideration of tagging behavior.

2.5 Motivation of tagging

Behind of tags, many researchers believed that taggers have motivations to give tags. To study of tagging motivations, researchers conducted qualitative interviews to taggers and they found two majors reason for tagging were archiving and sharing.

Hammond and his colleagues (2005) found that tagging motivations are different in various tagging systems. While many people tagged to share pictures with friends in Flickr, articles in Delicious were motivated for self archives. Nov, Naaman and Ye (2008) argued that the motivations might also diverse even in the same tagging system, because the motivations were divers among people. Besides, tagging motivations also influence on resulting tags. For archiving motivations, the tags are more biased on personal opinions and personal terms. On the other hand, taggers who are aware of sharing will have more popular tags which are well known by other people (*Strohmaier, M.,* Korner, *C. and Kern, R., 2010).* There are 2 factors which influence on taggers: personal tendency and community influence:

Firstly, Sen and the colleagues (2006) defined personal tendency as preferences, knowledge, experience and beliefs of topic and tag. Additionally, the topic interest and tagging familiarity affect to tags: "*the high familiarity with the concept of tagging, Web directories, and social tagging systems are significantly and positively associated with high tag effectiveness for content sharing*" (Lee, Goh, Khasfariyati and Chua, 2009, p.1). Stoyanovich, Yahia, Marlow and Yu (2008) have pointed out tagging also is influenced by user needs. For example, during the holidays, people tend to visit and search for travel sites more than others. Also, he has same findings of tagging which is influenced by community tag.

Secondly, community has strong influences on personal trends. Taggers may think other people have more knowledge of the content than themselves. Nowadays, many systems have suggested tags which based from communities' tags in order to support user does tags well.

Ames and the colleague (2007) have conducted a qualitative study about motivations of tagging at FLICKR and Zonetag. They interviewed and finalized 4 types of tagging motivations as follows:

- *Self organization*: Some taggers said they were organizers as well as they liked to order things. As the result, they could use tags for later retrieving.
- *Self communication*: some taggers intended to their personal information, quality of document, and feeling about a person in the contents for future recall depictions.
- *Social organization*: taggers wanted their tags (or photo) could be found by other members. However, for personal photos, taggers preferred to share with family and friends via email as well as other private ways than by tags.
- *Social communication:* taggers could add some specific tags for sharing with public or family and friends.

In the contrast, Nov, et al. (2008) proposed a new methodology to investigate the motivation by quantitative surveys. They stood on the scale of perceive tagging behaviors: Self (do tag for organizing and browsing myself), Family&Friends (do tag for organizing and browsing family and friends) and Public (do tag for organizing and browsing other users). However, the study was conducted by both quantitative and qualitative research because the authors used interview results to finalize the outcomes insides of the quantitative. Finally, they found taggers preferred to use tags as self organize and public sharing. Users from Flickr could share their tags for family and friends via other system like email (it is more private site). For this point of view, designing tasks for tagging, we should need to focus on self organization or retrieval and public sharing motivations.

Generally, many researchers try to answer why people do tag for personal and social reasons. It seems these studies focused the output of tags then figured out the drives behind tag output, as these papers used log analysis in both qualitatively and

quantitatively. Thus, there still needs to study about relationships between motivation, tags, behaviors as well as tagging systems, tagging objects, etc.

2.6 Cultures of tagging

Dong, et al. (2010) experimented on different image tagging groups: European Americans and Chinese. They had found that for the tags of European Americans were focused on the main objects more than the tags of Chinese. On the other hand, Chinese were more likely assigned to the overall description or relations between objects in the images.

Stoyanovich and the colleague (2008) found heterogeneous users brought about many kinds of tags. In addition, Chi Mytkowicz (2007) cited that different users used different terms to describe the same thing. This is an issue of spreading tags.

In regards of the various cultural tagging, some papers have been conducted to investigate multiple languages of tags. Tags are represented in a common language such as English, as the tags will depend on document language (Hammond, Hannay, Lund and Scott, (2005); Vuorikari, Ochoa and Duval, (2009). Moreover, Guy (2006) found in common tagging websites, tags are multiple languages because crowd-sourcing such as CiteULike, Deli.cio.us and CiteSeerX which are collaborative sites.

2.7 Types of tags

Tags can be categorized as following types: topic (what or who is it), time, location, type of document, authors/owner/creator, opinions/qualities, usage context (task organization) and self reference (Golder and Huberman, (2005); Firan, Nejdl and Paiu, (2008); and Sterken, (2008). Most of the authors used experiments and log analysis to figure those types. Korner (2009) has generated from the literatures of tag types that there are two kinds of taggers:

- Categorizers, who use tags to organize their resources for easier navigations, might be useful for us to summarize the resources.
- Describers, who want to accurate the tags for later searching or browsing, contribute for searching retrieval and knowledge acquisition.

In practice, some tags are complicated to be categorized such as "*vacation*" which could be topic or usage context. In addition, the types of tag depend on resources they are annotated so that the level details or types of tags could be different such as music or movie or image as well as text (Bischoff, K.; Firan, C. S.; Nejdl, W.; and Paiu, R., 2008, p. 212).

2.8 Semantics of tag

Vuorikari, et al. (2009) have conducted quasi-experiment of multiple and found that:

- Many tags include multiple terms, bundle together without comma separation.
- Taggers intend to use multiple languages more than tagging by their mother tongue. However, the finding could be right in case of high education users. They tend to tag by document languages.

Wu, et al. (2009) used visualization method to figure out some tagging structures:

- Tags with the same or similar meaning are picked by users' predilection. For example, "football"/"soccer" or "howto"/"tutorial" could be used by different users towards same thing in tagging.
- One person's tag usage might not be consistent, due to the arbitrariness in tagging. For example, at the same subject but they use different tags in different times.

There are three problems of semantic tags: polysemy, synonym and basic level variation. For example Spitieri (2007), Golder and the colleague (2005) showed that tags have these characteristics:

- Polysemy: Window may refer to a hole in the wall or to a pane of glass.
- Synonym: TV or television
- Basic level variation: dog or beagle

Through transaction log, Cattuto, Baldassarri, Servedio and Loreto (2007) found most of the tags in Delicious system are mainly nouns, synthetic descriptions and no grammatical structures. However, Spitieri (2007) has found that tags or folksonomies are grammatical structures by data log analysis from three social net work site (Delicious, Furl and Technorati).

According to Spitieri (2007) found that the tags are ambiguous and inconsistency. Firstly, tags are misspelled because taggers may forget right spellings. In this case, other users cannot search and connect to the misspell tags. Secondly, tags might have abbreviation, initialism and acronym. There are a number of tags which are labeled as abbreviation, initialism and acronym. For example, the SF might be Science Fiction or San Francisco. Finally, tags might have neologism, slang and jargon of tags. For example, instead of "police man", some Vietnamese call it as "Dove".

2.9 Solutions for tagging issues

Chi and Mytkowicz (2007) suggested model to support vocabulary issues in tagging. The idea is creating a function in which user can click on words of paragraphs to keep them as tags later. There are some other solutions to avoid vocabulary issues such as Lazy Sheep's model which auto-tag and auto-describes user's bookmark. However, these authors argued that it could lead taggers into cognitive barriers of tagging. To solve it, user communities' tags could be good solution to adjust this argument.

According to Ames et al. (2007), they suggested some solutions for tagging systems:

- Making tags pervasively and multi-functionally for tagging and searching in both mobile devices and desktops devices.
- Making tagging interface become easier for users, as the fact that tagging activities was increasing in such kind of good interface and tagging function such as FLICKR.
- Not forcing user to do tags because it effects to the quality of tags.
- Using relevant tags which need to be added caution for clarifying contents of tags to suggest for users. However, he did not show how to do it in realities.

Yin, Xue, Hong and Davison (2010) based on Bayesian approach which integrated three aspects to predict user's tags: ego-centric effect (user interest or user profile), environment and web page content effect. Through the experiment, the model can improve the system to suggest tags right person, right content. They also noticed that the tagging system weights

on user interests as well user profile. However, it still has issue that whether users give enough personal information for system or not.

There are some issues of tag structures in Spitieri (2007) and the author has suggested some ideas to integrated cataloging systems into tagging systems such as:

- A link to recognized online dictionary or thesauri as well as Wikipedia for user's decision of tagging.
- The cataloging system creates clear- written recommendations for choice and form of tags which include different type of semantic and components tags. LibraryThing could be such kind of this example, as they use communities' tags to recommend and use related subjects from traditional catalog to suggest tags for taggers.

SUMMARY

Simply, tagging is such kinds of user generated content, which uses to index objects. It also relates to some terms: user tags, folksonomy, personal ontology, and user contributed classification, free keywords, etc.

There are two major tagging systems: narrow and broad tagging systems. Flickr and LibraryThing is representative as an example. Different tagging system will drive to different behaviors

Almost studies about tagging behavior have been done by experimental methods. For some papers about motivation and cultures in tagging, they applied both qualitative and quantitative methodologies. It needed deep studies on tagging behaviors but process of tagging behavior is a shortcoming of this field. In the researcher' ability of searching and review, there is only on blog post about cognitive tagging process. However, the essay only raises an idea about tagging process but not for tagging behavior process so that we still need to do research on this.

CHAPTER 3: RESEARCH METHODOLOGY

This chapter presents usage of research methods to collect and analyze data for answering the research questions. In order do that, data collection methods such as sampling techniques, designs for data collections and process of data collection; and data analysis methods also will be discussed.

3.1 Research methodologies

As tagging behavior process relates to cognition and behaviors, the study is going to base on qualitative research to answer the research questions. In order to study on tagging behavior process, there are some methods to investigate it such as observation (observe tagger in realities), experiment, questionnaire survey and interview, etc.

On one hand, there are some quantitative approaches to study about tag types, semantics of tag. Guy (2006) studied about tagging linguistics by log analysis; Golder and Huberman (2005); Firan, Nejdl and Paiu (2008); Sterken (2008) collected tags from the quantitative experiments to visualize tags to define a variety of tag types. On the other hand, tagging motivations, tagging cognitions are studied by qualitative. Rashmisinha (2005) based on his observations and experience of cognitive theories and tagging to figure out cognitive tagging process; and Ames and Naaman (2007) have conducted a study about tagging motivations by interview taggers.

According to many papers of cognitive psychologies, most of them used think aloud experiments (talk aloud or talk-louder) to investigate user cognitive, feeling as well as process of decision making (Van Someren, M. W., Barnard, Y. F., and Sandberg, J. A. C., 1994). In this study, talk aloud experiments will be the main method to collect data. The talk aloud experiment expects informants to say anything goes through their minds. Doing it, researchers might know their cognition, feeling and behaviors as well. Besides, the study also uses observation, questionnaire survey before test and interview after that.

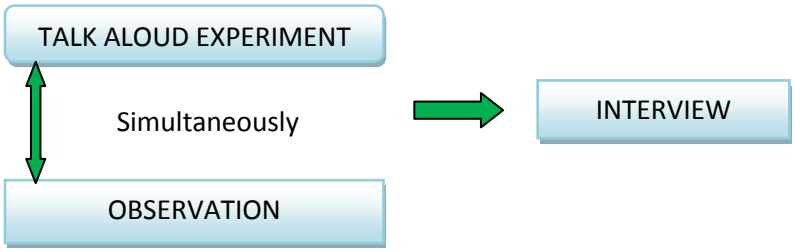

Figure 3.1: Methods of data collection process

3.2 CiteULike, LibraryThing and Flickr, YouTube as case studies

There have many tagging sites to support user organize their collections, browser and sharing. This study will choose CiteULike, LibraryThing and YouTube or Flickr as experimental tagging sites because of different types of documents such as book, articles and video and image.

In the further experiments, informants will tag for articles and fictional books in CiteULike and LibraryThing. For video and image, informants will search and read in YouTube and Flickr but they will tag in papers. Thus, the basic steps to tag in CiteULike, LibraryThing as 3 figures as follows:

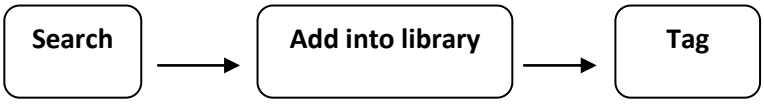

Figure 3.2: Quick tagging steps in CiteULike

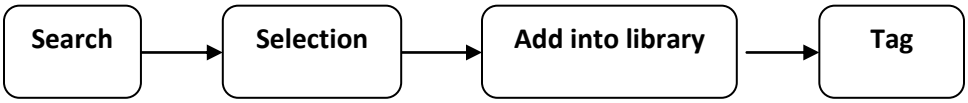

Figure 3.3: Normal tagging steps in CiteULike

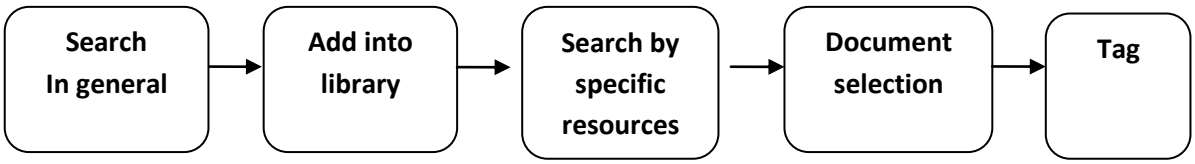

Figure 3.4: Tagging steps in LibraryThing

3.3 Sampling technique

3.3.1 Size of samples

According to some papers about self-talk experimental method, the number of informants is from 5 to 15. Nielsen (2000) argued that from informant 7 or 8, he find coincidental data to test the website usability. As the time limitation, this study will have 3 pilot informants and 9 official informants to be tested.

3.3.2 Criteria for recruitment

- *Multi-education backgrounds:* Including students, teachers, librarians as well as other staffs in a variety of areas.
- *Tagging experience*: As study of tagging behavior process means to observe how taggers behave in tagging site, taggers who have fair to excellent tagging experiences are major samples.
- *Multi-cultures*: From DILL course, it is possible to have many people from many countries such as: China, USA, Colombia, Venezuela, Denmark, Romania, Germany, Netherland and Uganda.
- *Language for experiment*: English is chosen language for testing since topic reading, tagging systems and talk-louder in English. Therefore, choosing international student is feasible, since they are studying in English so that they can have more English communicative skills.

3.4 Data collection

3.4.1 Objectives

The first objective is investigation in different processes among different informants and different types of objects and sites. Secondly, the study tries to understand actions, behaviors, cognitions and factors influence on tagging process. The final are observations of tagging issues during tagging behavior process.

3.4.2 Tasks designs

Tagger may choose their interesting topics for searching, selecting and tagging. Since in realities, taggers may tag for topics or documents that they really want to add into their libraries. As the results, before experiment, the study has a preparation for informants which includes instructions and requires them in choosing topics for the tests.

Text documents such as article, book and video, image are main objects for tagging experiments. Articles and books from CiteULike while images from Flickr and video from YouTube. For time consuming in up loading video and image, searching and exploration are done by Flickr and YouTube but tagging will be conducted in paper. Also dealing with up loading issue, informants are expected to tag in paper for tagging full text.

There are three major types of topics: the first is factual topics such as articles and full text tagging about sex safety topic; the second is fictional topics which there are five famous fictional options will be chosen by informants; the final is entertainment topics such as video and image. Doing a variety of topic, the study can compare how taggers think, feel and behave among different topics and tagging objects.

Tagging tasks design:

Task 1: Tagging in CiteULike

From LibraryThing, choose and search a topic you like and then collect one possible document at the same with that topic. Then give tags for chosen documents, and while you do so, try to say everything that goes through your mind.

Task 2: Tagging in a paper

Context:

You are working in a Sex Safety Project which aims to help the youths become aware of how to protect their sexual health. Your responsibility in that project is collecting related documents.

Responsibilities:

There is an article "Sex safety: the best protection against sexually transmissible infections, including HIV/AIDS" in the laptop desktop.

Please do anything you wish (you can use internet to check information if you want) to tag it in a paper and while you do so, try to say everything that goes through your mind.

Task 3: Tagging in LibraryThing

Please choose one of the books following the table:

Book	***Title 1***	***Title 2***	***Author***
1	*Harry Potter and the Philosopher's Stone*	*Harry Potter and the Sorcerer's Stone*	*J. K. Rowling*
2	*Where Rainbows End*	*Rosie Dunne*	*Cecelia Ahern*
3	*The Iron Man*	*The Iron Giant*	*Ted Hughes*
4	*Little Mexican*	*Young Archimedes*	*Aldous Huxley*
5	*Miss Smilla's Feeling for Snow*	*Smilla's Sense of Snow*	*Peter Høeg*
6	*Northern Lights*	*The Golden Compass*	*Philip Pullman*
7	*The War of Dreams*	*The Infernal Desire Machines of Doctor Hoffman*	*Angela Carter*

Then, search for both titles of that book and give tags for them. When doing the task, please try to say anything that goes through your mind.

Task 4: Video tagging in YouTube

Pick any video on YouTube of your choice and tag it in paper, while doing so; try to say anything that goes through your mind.

Task 5: Image tagging in Flickr

Pick any picture on Flickr of your choice and tag it in paper, while doing so; try to say anything that goes through your mind.

3.4.3 Instructions

The instructions help informants how to search, to read and to tag in concrete tagging system. Furthermore, they also explain about how to do the test and some notices of talk aloud and questions during testing. (See also appendix 2)

Structure of the experiment:

- Time to take test: approximately 1 and an half hours
- Process of test:
  - Pre-test : 5 survey questions
  - During-experiment : 5 tasks
  - Post-experiment : 5-8 interview questions.

3.4.4 Instruments

Room for experiment requires:

- Computer, tape recorder, internet for informant can search, check and do any works to support experiment
- Quiet and no disturbed room for informants can think aloud.
- Friendly room for them to do tag naturally.

Software to record the experiment is Depute Video Capture Screen which is open source software and has some basic functions as follows:

- Capture all informants' activities on the computer screen.
- Record the sound from informant's doing and talking. However, to be sure of the sound quality, I am going to use tape recorder for both talk aloud experiment and interview.
- Length of record is unlimited.
- The software will be run for each session to decrease the size of file.

3.4.5 Moderator

Moderator should follow these instructions for the testing:

- Keep neutral body languages and friendly to make informant do tag as nature as possible.
- Be encouraging and know how to warm up talking louder from informant when they feel tired along the long test. For example, after each task, moderator let informant

have some minutes to relax as well as ask the sentence before each task start "Now it is time for next task, please keep on saying anything during this task".

- Avoid to ask or disturb informants when they are talking, because this behavior can stop their feeling or thought. So, just ask when the talk is ended.
- Keep informants talking louder when they stop talking (they may be tired or they forget to talk) by some questions: "What are you thinking now? Please say anything while you are doing it".
- Be naturally inquisitive to ask informants when they are stuck in somewhere, moderator can ask "what are you thinking about this?", then they may response and moderator will take note down in order to interview them later.
- In case informants completely give up the task, it is possible to allow them change to next tasks, as if moderator also cannot help them anymore to be positive results.
- Let informants feel free and comfortable to do tags. As the result, moderator should seat far enough from their place to observe and take note.

3.4.6 Process of experiment

***Pre-experiment*:** Send the preparation of experiment to informants which introduce of tasks and suggest them to prepare interesting topics for each of tasks. Besides, in the test day, informants are expected to fill up 6 survey questionnaires about rating of tagging experience and topic experience.

***During-experiment*:** Informants will follow instructions to finish 5 tasks by talk louder, while observer note down questions for interviews or answer for informant's questions.

***After-experiment*:** Informants will be interviewed in 5 or 8 questions.

3.4.7 Observation design

***Objectives*:** The observation is going to support for talk louder experiment because there will be right time and right things to interfere, answer and give questions for informants. Secondly, there will be additional questions which appear while testing will be prepared for interview.

Things to be observed:

- Questions from informants
- Encourage informants talk louder
- Strange tagging behaviors and cognition tagging

3.4.8 Interview design

Objectives: The aims are to know cognitions and feeling of informants during experiments. Moreover, the interview supports to know unclear voices, sentences, ideas, feeling and behaviors in the experiment.

Things to be questioned: (see also appendix 6)

- Feeling and decision making, motivations of tags
- Evaluation of tagging systems and suggestions for tagging system

3.5 Pilot study

There are 3 informants involved in the pilot study at the beginning of research. After the pilot, there are 3 major lessons for the official experiments. Firstly, the pilot recognized indexing skills of librarian influenced on tagging behaviors so that the informant's recruitments will be expanded to out of DILL samples. Secondly, however, librarian or indexing skills is the not major influences on tagging behaviors. For example, #P still met many troubles in tagging as she doesn't have tagging skill, even though she is a librarian and have indexing skills. The observation of #P shows that in this case, tagging experience is more important in tagging behavior. Thirdly, the topics in tasks should be chosen by informants more than compulsory topics.

3.6 Data analysis methods

The data gathered from talk aloud experiments and open interviews are mainly qualitative. Thus, a method called constant comparative analysis is applied for coding and categorizing data. This method comprises of three steps including coding, categorizing and clustering. Since the data is gathered from experiment, there is an additional step is transcription.

For transcription, as tagging behavior processes characteristics (study about cognitions and behaviors), it is necessary to type words down as verbatim as possible or even silences from informants. Nevertheless, for interview, main ideas will be transcribed.

For coding, there will be two main stages such as raw data grouping and coding. Firstly, it is important to recognize raw data which have the same meaning in a group which will be coded by numbers. After that, each group will be attributed a code which represents a theme that data is associated with. The code is identified by name of behaviors, cognitions, and tagging process.

For categorization, similar themes or common opinions are merged together to form categories. Simultaneously, data placed under each code is also joined together.

Finally, these categories are clustered around each research question to identify which categories could be answer for research issues. Some categories may be related to more than one research question. If categories do not fit to any research issues in the study, it might be used for further research recommendation.

3.7 Ethical consideration

The gathered data of 9 informants has been illustrated for the data analysis to understand about process of tagging behaviors. The thesis will not evaluate or ranking of responders. Moreover, names and other personal information will be coded and hidden to keep privacy rights for the informants.

CHAPTER 4: DATA ANALYSIS

Abstract:

This chapter presents the process of tagging behavior and the tagging challenges. Firstly, the process of tagging behavior are discussed in several aspects such as: tagging experience and perceptions, classification of taggers, explorations, tagging decisions, tagging motivations and tag formulation, factors influence on tagging behaviors and types of tags. Secondly, 4 main tagging challenges which relate to information resources, tagging functions, vocabularies issues and patient issue are presented.

4.1 Geographic of informants

4.1.1 Tagging perception and tagging experience

There are 9 informants participated in the experiments. Most of the informants were aware of the concept of tag. They stated that tags can be keywords which appear frequently in the text or represent for documents.

Informants	Excellent	Good	Fair	Non-skill	Experienced sites
#W				✓	My Blog, YouTube, Tumsr
#R			✓		Facebook
#A			✓		Facebook, Picassa, Flickr
#K			✓		Facebook
#C			✓		Facebook
#L		✓			Facebook
#Jen		✓			Facebook
#E		✓			Facebook, Picasa
#J	✓				Flickr, Youtube

Table 4.1: Tagging experience

Informants	Background	Experience
#W	None librarian	None
#R	Librarian and studying tagging	Fair
#A	Librarian	Fair
#K	None librarian	Fair
#C	Non librarian	Fair
#L	Librarian and studying in tagging	Good
#Jen	Librarian	Good
#E	Librarian	Good
#J	Librarian	Excellent

Table 4.2: Informants' background

Firstly, in the table 4.1, #R, #A, #K and #C claimed that they have fair tagging experience while #L, #Jen and # E thought that they have good tagging experience. Besides, #J claimed he has excellent tagging experience. The table 4.2 show that #W, #K and #C are none librarians whereas #R, #A, #C, #Jen, #L and #J are librarians

Secondly, the librarian skills influence on tagging behaviors which will be illustrated by the table 4.3 below. Librarian informants rated in scales of tagging experience from fair to excellent, whereas none librarians had tagging experience from fair to none. Additionally, #L and #R have very good knowledge on tagging which are their topic interests to study. In the experiments, #L and #R showed that they are strongly affected by librarian skills:

- Indexing skills influenced on tagging behavior: "*Because I am librarian so that I know it should not have too many tags for a document as it could make spreading of document*" (#R).
- Compound tagging experience influence on tagging behavior: "*I need to come to tag clouds in LibraryThing to see how other taggers tag for a book made into film*" (#L).

Tagging experience	Excellent	Good	Fair	No
Librarian	1	3	2	
None-librarian			2	1

Table 4.3: Compare of tagging experience and informant's background

Finally, most of the informants are real taggers within image and video tagging systems. Facebook is a well-known tagging site for the informants. Besides Facebook, some

informants such as #J, #A and #E have experiences on YouTube, Flickr and Picasa. Most of the informants preferred to tag with videos and images for the following reasons:

- User need of video and image. "*I need to organize my pictures because I have a number of pictures. Can you image how I retrieve those pictures without good tags?*" (#C)
- The video and image are entertainments. *"To tag for text which is more academic such as safety sex for youths, you need to be careful than tag for videos or image*" (#R)
- The visibility of video and image. "*The video and image are visible and easily to understand. Moreover, they are also exciting to watch and see for tagging*". (#Carl)

4.1.2 Classification of taggers

Throughout the data, there will be two main kinds of taggers: confident taggers and certainty taggers. Two of them will be defined in the paragraphs below. The aim to present 2 types of taggers is helping readers in following of the chapter data analysis easily.

Tasks	**Confident taggers**	**Certainty taggers**
Task 1: Tagging in CiteULike	#R, #Jen, #J, **#E**	#A, #L, #C, #K, #W
Task 3: Tagging in LibraryThing	#R, #Jen, #J,	#A, #L, #C, #K, #W and **#E**
Task 4: Tagging in YouTube	#R, #Jen, #J, #E and #A, #C, #K, #W, #L	
Task 5: Tagging in Flickr	#R, #Jen, #E and #A, #C, #K, #W, #L	
Task 2: Tagging with full text		#R, #Jen, #J, #E and #A, #L, #C, #K, #W

Table 4.4: Classification of taggers

The first is **confident taggers** (#R, #Jen and #J). The confident taggers based on topic experience to tag without information seeking before tagging. For example, #R tagged without reading content of the article about videogames, which he has never read before. The confident taggers follow the tagging process as the figure:

Stage 1: Topic experience		**Stage 2: Tagging decision**			
Scanning	Memorization	Prediction	Motivation	Tag formulation	Consideration

Figure 4.5: Tagging process of confident taggers

The second is **certainty taggers** (#L, #E, #C, #K, #W and #A). The certainty taggers based on topic experience and information seeking to tag. For example, #A has good knowledge of Brasov history but he still wanted to read the abstract of the article to tag. The certainty taggers have tagging process as the following figure:

Stage 1: Topic experience	**Stage 2: Exploration**	**Stage 3: Tagging decision**			
Memorization	Reading, understanding	Prediction	Motivation	Tag formulation	consideration

Figure 4.6: Tagging process of certainty taggers

For YouTube and Flickr, most of the informants are confident taggers since they can choose images and videos which they knew before tagging. For the full text, since they were required to read or scan the booklet of Sex Safety, most of them read and explore the text as certainty taggers. As the results, my study only classify taggers through Task1 (Tagging in CiteULike) and Task3 (Tagging in LibraryThing). In the Task1, most of the selected documents are not read before tagging. In the Task3, some of the selected books are not read before; while the others are read before (those documents will not be examined for the certainty taggers).

In the table 4.4 of the Task1 and Task3, it is clear that #R, #Jen and #L are confident taggers while #L, #A, #C, #K and #W are certainty taggers. Specially, in the Task1, #E is considered as confident tagger whereas in the Task3 she is certainty tagger. This happen because in the Task1 #E chose a document which she has read before. In conclusion, #E is included in the certainty taggers.

4. 2 Explorations

The exploration stage will be divided into 2 main kinds: text exploration and video, image exploration. The confident taggers and certainty taggers in the text exploration will be

discussed separately. On the other hand, the video and image exploration will be analyzed on the entire informants, since there is no distinguish between 2 types of taggers in the video and image tagging.

4.2.1 Exploration of confident taggers

It doesn't matter whether they had read documents or not before, the exploration of confident taggers has two stages: scanning and memorization.

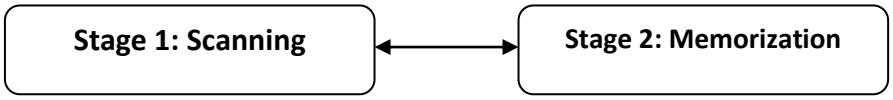

Figure 4.7: Process of read document exploration

Firstly, taggers scan main features such as titles and authors in the tagging pages to confirm the relevant documents. The scanning was ended quickly when taggers saw necessary information such as titles or familiar terms they have known before. On the other hand, if taggers could not find the information they want to see, they might carefully look at abstracts or other metadata.

Simultaneously with scanning, taggers might memorize from topic experience what they have known. The memorization stage is appeared at both confident taggers and certainty taggers. The memorization will help taggers to narrow the topic of documents into a topic experience memorably and understandably: "*I need to put the document into its topic experience such as the "sex safety" will be put into "epidemiology" to easily memorize and understand"* (#L). Moreover, taggers also memorized pictures, facts, names of stories, names of authors, main people in the books as well as feelings about the books. There are some characteristics in the memorization:

- Some informants easily remembered popular and general information than detailed information. The table 4.8 shows the popular and general tags by the coincidental tags of the book "*Harry Potter and the philosopher's stone*":

Harry Potter and the philosopher's stone	#Jen	**school, wizards, witches, magic, children, fantasy, fiction**
	#E	Professor **Dumble Doe, Hogwarts School, Harry Porter, Lord Voldemot, Ron Weasley, Wizards, J.K.Rowlling**, British fiction
	#K	**harry potter, j.k.rowling, Hogwarts, dumbledore, voldemort**, philosopher stone, **ron weasley, hermioni granger, children**, adults, the deadly hollows, **fantasy, magic, j.k.rowling**, Volume

Table 4.8: Tags of the "Harry Potter and The philosopher's stone"

- The tags might be influenced by level of tagger's memory. For example, #R remembered many details of the stories, while #E just listed four stories in the Bible book which she read long time ago.

4.2.2 Exploration of certainty taggers

Comparing to the confident taggers, the exploration of certainty taggers are more complex through searching, reading and understanding document basing on their knowledge and the document explorations.

4.2.2.1 Explorations within the tagging pages

The titles, abstracts, full text, table of content and tag clouds are mostly looked and scanned in CiteULike and LibraryThing. Comparing between two groups, the librarian informants and the non librarian informants, there will be various levels of explorations in the illustrated table 4.9 below:

Tagging with	Librarian informants	Non librarian informants
Articles in CiteULike (no full text)	Titles, abstracts, **authors, year publishing, suggested tags, etc**.	Titles and abstracts.
Books in LibraryThing (no full text)	Titles, descriptions, **tag clouds, links to full texts, links to cataloging sites.**	Titles and descriptions.
Full text tagging	Titles, table of content, headlines, detailed information, terms extractions, etc.	Titles, table of content, headlines, detailed information **(lesser than librarian informants**), terms extractions (**lesser than librarian informants)**, etc.

Table 4.9: Exploration of the librarians and non librarians

In the table 4.9, the exploration of non librarian informants were simpler than librarian informants. Additionally, not at all librarian informants had sophisticated exploration and not at all non librarian informants had simple exploration. For example, although #A is librarian, she simply looked at title and abstract to tag for the article of Brasov history. Nevertheless, #C is non librarian but she had sophisticated exploration through the titles, table of content, headlines, detailed information and terms extractions of full text tagging. In general, the exploration has some characteristics below:

Firstly, most of the informants wanted to get general understandings of topic through metadata such as titles, abstracts, authors and year publishing.

Secondly, most of the informants looked at abstracts, full text and some web pages which provide more information about the documents. For instance, #E came to Wikipedia to read abstract about "*Harry Potter and the Philosopher's Stone*"; #Jen came to a specific website which included abstract of the "Safety Sex" to read for tagging.

- Normally, people tried to read the titles and abstracts of articles and books. Many informants were satisfied with the abstracts in the CiteULike. Since the experiments showed that from the abstract, most of the informants could make tagging decisions without struggles in exploration. For example, #K said in tagging with CiteULike: "*My tags are mainly based on the abstracts*".

- There are many ways to read abstract. Some informants read carefully to understand the articles or books, while others informants just scanned some important terms in the abstracts then extracted or copied them into tag cloud. For example, #L, #C, #K and #Carl tried copying some important terms in the abstracts and full text of the articles into Microsoft Word or taking notes in the paper for tags later.
- Furthermore, the way to present information or abstracts influenced on tagger's feeling. Most of the informants said in the interview that the information structure of LibraryThing is not good for them to look and tag easily: "*LibraryThing should put abstract directly under title so that people easily to see and read it, since LibraryThing has complicated information structure*" (#C). (See the figure 4.10 and 4.11)

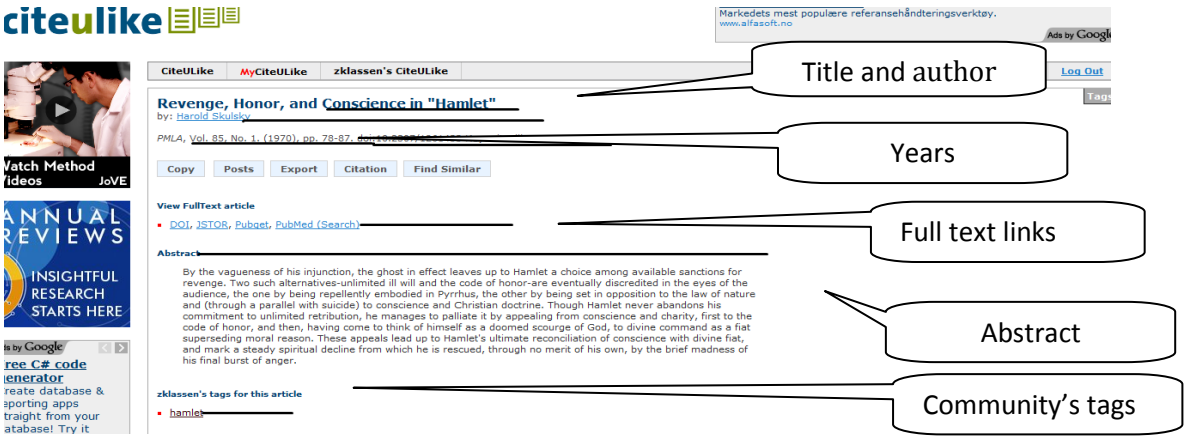

Figure 4.10: Information structure in CiteULike

Figure 4.11: Information structure in LibraryThing

The third level is looking at subject headings from libraries or communities' tags. At this level, only #L who is a librarian and has interests to study of tagging was looking at the tag clouds in LibraryThing and subject headings from WorldCat to make references for tagging decisions. For example, #L came to the tag clouds in LibraryThing to see how other taggers tagged for the book "*Harry Potter and the Philosopher's Stone*" and tried to look at the library's subject headings of the book to tag. On the other hand, the other informants didn't want to reuse the suggested tags in either LibraryThing or CiteULike: "*Oh, it not good for my tags*" (#R); "*I don't think I will use these tags because they are not fit to my idea and I want to give my own tags*" (#Carl). Furthermore, #C, #K and #W who are non librarians were not attractive by the community's tags.

Finally, comparing with the guidelines of subject indexing in the DDC[1] and the ISO[2], the exploration of the informants is simpler than the real indexers. The guidelines suggested several things to look for indexing such as title, table of content, the preface or introduction, a scan of text, bibliographical references, cataloging copy from the centralized cataloging services or original cataloging, as well as the other outside sources such as reviews, reference works, author's keywords (Mitchell, 1996, xxxv; ISO, 1985, 5.2). The informants normally looked at some part of them such as titles, abstract, table of contents, scan of text and subject heading or suggested tags.

In conclusion, titles and abstracts are the most popular things looked at by taggers. As the result, the quality of presentation of these factors affects strongly to result of tagging.

4.2.2.2 Explorations outside of the tagging pages

Information seeking outside of the tagging sites happens under the following conditions:

- When informants needed to understand topics but abstract from LibraryThing or CiteULike was short. For example, #K wasn't satisfied with the descriptions (abstract) about Harry Potter in LibraryThing thus she came to Google books to

[1] DDC is Dewey decimal classification.
[2] ISO is International Organization for Standardization

read or "*I cannot find the difference between 2 titles of the book "Harry Potter and the Philosopher's Stone" from the review. I think I should come to Google to check*" (#E).

- When informants doubted and considered about facts from the abstracts and the text and there is no place in the tagging pages to correct the doubt and consideration: "*I am wondering do the Inuit is right with a place in Greenland. I will come to Wikipedia to check it, yep*" (#R).
- When informants wanted to know meaning of terms from text. This phenomenon happened in two cases. Firstly, as non English natives, the informants might not know the meaning of the words. For instance, #C came to a dictionary to check "Chinese Lantern". Secondly, taggers wanted to know meaning of specific terms such as "*Trichomonas", "Herpes", "Genital wants", "Chlamydia", "Gonorrhoea*" which was searched in the Wikipedia to check scientific meanings by #R.

4.2.2.3 Topic memorization

Compared to memorization basing only on topic experience, the topic memorizations within reading and understanding are complex and have some characteristics as follows:

- The memories oriented for taggers in reading and understanding the documents: "*I need to put the document into its topic experience such as "sex safety" will be put into "epidemiology" to easily memorize and understand*" (#L).
- The memories gave taggers some particular terms to be explored in the further seeking: *"I want to find a name of project which I have known before*" (#L).
- The more #L read abstracts or full text, the more she could memorize terms and topics. As the result, the memories and explorations interact together and help taggers have better understandings to tag.

4.2.3 Explorations of video and image

Compared with textual documents, video and image has different ways of exploration. In this part, memorization and doubt will be discussed.

4.2.3.1 Video and image memorization

Because most of the informants could choose videos and images which they like to tag, the videos and images were familiar documents with them. Watching videos and images reminds taggers about content of the videos and images. For example, #E watched video of Miley which reminded #E about the name of singer, name of the song as well as other details which appeared in the video.

4.2.3.2 Doubts about video and image

Some informants doubted about detailed information in the video and image. For example, #J wanted to know genre of the music as he wanted to tag for music genre. However, he couldn't find any useful information from YouTube because of lacking the description for this video. Besides, #Carl wanted to find names of places, people in the video to tag but she couldn't do it, since there was no description in the video.

4.2.3.3 Summaries of exploration stage

To conclude for the exploration stage, there will be three approaches to explore documents before tagging: exploration of confident taggers, exploration of certainty taggers and exploration of video and image. In exploration of confident taggers, taggers will spend on two steps, scanning and memorization. In exploration of certainty taggers, tagger will have two basic methods to explore, explore within the tagging pages and outside of the tagging pages. For both methods, taggers will memorize what they have known about the topic of documents and then try to read and understand adequately for tagging. Finally, the video and image exploration is simpler than the other approaches.

4.3. Tagging decision

The decisions making of tagging are defined as "*Choosing the best category is something we do all the time. We see an animal – it could be a dog or wolf. We make a quick judgment. This is a basic cognitive process – putting things into categories*" (Rashmisinha, 2005, p.2). The

definition means cognitive tagging decision is a process of categorization and selection for the best category of documents.

From the experiments, there are two formulations of tagging decision, tagging decision of the confident taggers and certainty taggers:

Tagging decision of the confident taggers:

Choice			**Adjustment**	
Topic experience	**Prediction**	**Motivations**	**Tag formulations**	**Considerations**
- Memories - Beliefs	Searching phrases	- Self archives - Self communication - Social archives - Social communication	Extraction of Memories, motivations	- Number of tags - Correction of tags

Table 4.12: Tagging decision of the confident taggers

Tagging decision of the certainty taggers:

Choice			**Adjustment**	
Topic experience and document exploration	**Prediction**	**Motivation**	**Tag formulation**	**Considerations**
- Memories - Beliefs - New knowledge of documents	- Phrases - Orientation for exploratio n	- Self archives - Self communication - Social archives - Social communication	Extraction of memories, motivations, document exploration	- Number of tags - Correction of tags

Table 4.13: Tagging decision of the certainty taggers

After the exploration stage, taggers are distinguished clearly into two groups. The first group based on only personal experience and belief and the second group based on personal experience, belief and document exploration. This stage leads to differences of tagging decisions through two formulations of tagging decision in the table 4.12 and 4.13.

Consequently, there will be two main stages of the tagging decisions, choice and adjustment.

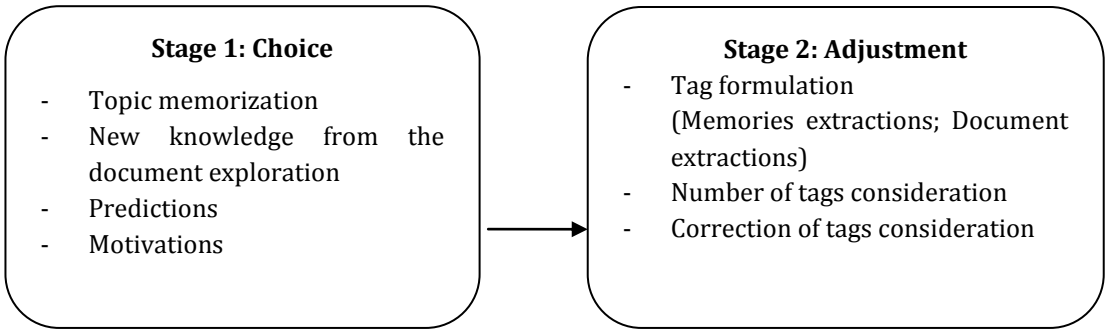

Table 4.14: Tagging decision stages

4.3.1 Tagging prediction through the match between tags and searching phrases

Seemly, most of the informants have predictions before tagging decision stage. The predictions can be seen clearly in the match of searching phrases and tags. Before the exploration, taggers searched with some searching phrases which can be any strings. Most of the searching phrases have three main levels of matching with tags:

1) Perfect match. For example, #R searched by "videogames" and the searching phrase is appeared accurately in the tag "videogames".
2) Partially match. For example, #R searched with searching phrases "*han havde været træls*" but only "*træls*" was appeared in the tags.
3) No match. For example, #A searched by "*The Golden Compass*" and assigned only one tag "*adventure*".

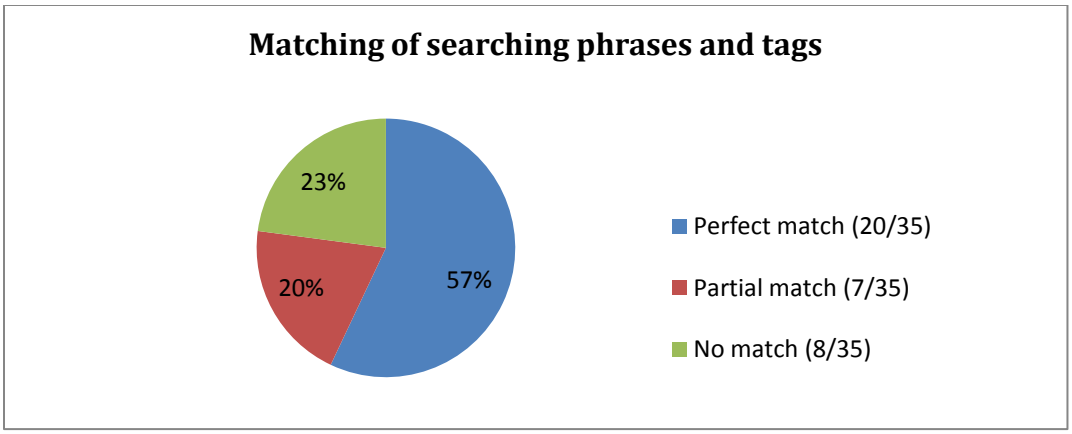

Figure 4.15: Matching of searching phrases and tags

In the table 4.15**,** there are relationships of the searching phrases and tags. The number of searching phrases have been matched exactly with the tags is 57% and the searching phrases also appeared partially in the tags with 20% while there are 23% of searching phrases are no match with the tags. Some observations aim to explain for the partial matches and no matches:

Firstly, most of the partial matches have the following characteristics:

- The searching phrases are titles but the tags are taken as a part of the titles. For example, #E searched by "*harry potter and the philosopher"* stone" and the tag is "*Harry Potter*" which is popular by readers.
- The searching phrase is a long sentence but the tags might be a specific term in that sentence. For example, #K searched by "*communication theories"* then she tagged "*communication*" which is the core meaning of the phrase.

Secondly, most of the no matches have the characteristics as follows:

- The searching phrases are general topic while the tags are more detailed. For example, #E searched with "*Bibles stories*" then tagged for some particular names of the stories such as *"Moses and the ten commandments, the birth of Jesus, the story of Joseph in Egypt, Noah ark".*
- The searching phrases are in detail but the tags are general. For example, #A searched with "*The golden compass*" but then tagged with "*adventure*".

As the result, even though the searching phrases are matched partially or not matched totally in the tags, there still have the relationships between them. The relationships are general to detail; detail to general; and partially. Firstly, in the no matches, the tags might be specific terms while the searching phrases might be general terms. Conversely, the relationship will be exchanged: the tag is general but the searching phrase is specific. Secondly, in the partial matches, the tags might be extracted from a part of the searching phrases.

4.3.2 Tagging motivations

According to the experiments, most of the informants have the same types of motivation discussed in the theories of tagging motivation. Ames and Naaman (2007) found there were four types of motivations, self organization, self communication, social organization and social communication:

Self organization: From the interviews, most of the informants said they strongly wanted to tag for their archives. For picture tagging, #C said "… *I want to tag to organize my huge number of pictures…You can imagine that I have bad tags, how could I find a picture in thousand of pictures*"

Self communication: some taggers intended to add personal information, quality of document, and feelings about a person in the contents for future recall descriptions. For example, #R tagged "to-do-list" to *"…remind me to the priority of travel plan that I will visit to the Northkap in near future*"; #R tagged for the "Miss Smilla feeling for snow" with tag "*wrong author*" because "*Another people did wrong author of the books so that I need to notice about it*" (#R).

Social organization: Taggers wanted their tags can be found by other members. For example, #L, #R, #E, #K and #J tried to have some popular terms in tags which other people may search by. Additionally, #L and #C used some explanations for the abbreviations as well as use synonyms for tags to be findable by other people.

Social communication: Taggers can add some specific tags to share with public or family and friends. The tag "*this-book-used-for-kid*" is an example. However, they are careful with sensitive and privacy things. In the study, this kind of motivation was not found.

Throughout the experiments and interviews about tagging motivations, most of the informants have common motivations: self organization, self communication and social organization. Firstly, most of the informants just tagged naturally on what they have known about the topic and what they understood about the documents. Secondly, some informants assigned tags which aimed to communicate with them self. Finally, there are a few informants who wanted to tag in a way that increased findability to other taggers. For example, #L tried to explained the abbreviation LAM (Libraries, Archives, Museums); #C tried to use synonyms for "kid" and "children" in the Harry Potter tagging. (#L is a librarian but #C is not a librarian)

For factual topics, almost all informants wanted to tag for self archives and social organization, whereas, informants more biased on self communications with fictional topics and entertainment topics. For example, most of the informants had more tags about personal opinions and feelings on topics.

4.3.3 Tag formulation

4.3.3.1 Tag formulation of confident taggers

In the adjustment stage of tagging decision, the confident taggers, who tagged documents without reading, have some extracted characteristics as follows:

- Firstly, the confident taggers met difficulties to recall the tags. For example, #E assigned tags for the book "*Harry Potter and the philosopher's stone*" with "Dumble Doe" instead of Dumbledore" and "Lord Voldemot" instead of "Lord Voldemort".
- Secondly, the tags of the confident taggers were more dependent on personal languages, experience and knowledge.

4.3.3.2 Tag formulation of certainty taggers

The topic experience helped the taggers to have ideas and directions for reading within the abstracts or full texts. For example, before reading the booklet "*Sex safety*", every informant read the titles before the reading. Most of them claimed that the title helps to define the main topic then they could easy to read later on. Additionally, the taggers based on text to tag so that there was less difficulty in misspellings. Since, when they didn't remember exactly, they can look from the abstracts, title, tag clouds or full text to copy exact the terms they wanted. For example, #C copied some terms from the description of "*Harry Potter and the Philosopher's Stone"* into her tags; #L came to tag clouds to see other taggers assign for the book transformed into film.

The exploration helps tagger in getting new knowledge and tagging idea. For example, A# has known about the book "*Northern Light*" which has a different name, "*The Golden Compass*". She believed that those names do not refer to the same book. However, during looking information from LibraryThing, she only found that they are published in different periods and the "Northern Light" is a dark material of Phillip Puma. As the result, she noticed it into her tag.

Comparing to the confident taggers, the certainty taggers had much more consideration and doubt about numbers of tags, corrections of tags and sharing motivations. When they couldn't know exactly the correction of considerations, most of them came to Wikipedia, Google, dictionaries and full text as well as tag cloud to check it. For example:

- #E in the fictional tagging, she wondered number of tags in the talk aloud.
- #C who is not a librarian tried to have synonyms "*kids*" for "*children*" tag as she thought other people will search by kids instead of children.
- #L tried to explain some abbreviations for other people can search and understand.
- #L came to full text and tag clouds to extract tags for fictional tagging in LibraryThing.

The certainty taggers more motivated to tag for social organization purpose than the confident taggers who were biased on self communication and self archives. During

tagging, while most of the confident taggers mainly tagged to organize documents, some certainty taggers considered to use synonyms and abbreviations for searching by others. Fox instance, #L assigned *"(libraries, archives, museums) LAM"*; #C assigned *"kids, children"*.

Comparing to the confident taggers, the tags of certainty taggers have these characteristics:

- Number of tags: According to the analysis above the confident taggers usually assigned averagely 3 or 4 tags for each documents while certainty taggers have from 6 to 8 tags for each document.
- More detailed with diverse types of tags such as topics, title, authors, publishing year, name of places and type of documents.
- Less tags in regard of personal opinions.
- The tags are more dependent on text.

Comparing to video and image, the tag extractions of text are more difficult. #C and #K said they felt easier to remember tags within video and image than text because the video and image are visible than text. Moreover, from the observations, most of the informants assigned tags for video and image in convenient way. However, this phenomenon happened because all the informants are not real taggers with text tagging so that they might feel inconvenience.

4.3.3.3 Tag formulation of video and image

Comparing to text tagging, video and image tagging have several different characteristics in tag formulations:

- Taggers felt more comfortable to tag because they were both watching and enjoying video and image. Moreover, they said the video and image are very visible so that they are easier to understand.
- Less consideration and doubt on tags. They said in the talk aloud and interviews that what they see from them is what they get understand.
- Since there was less text to describe about video and image, when taggers wanted to know name of places and time in the pictures or videos, they didn't know how to explore them. For example, #Carl wanted to know name of the library, place and

people which appeared in the video about mobile library services but she didn't know how to look such kind of information.

4.3.4 Tagging consideration

As the presentation of tagging decision, in the stage of adjustment, taggers will have some consideration about their decisions of tags. After having tags from tag formulation stages, some informants considered about number of tags, correction of tags and form of tags.

Firstly, there is much consideration about form of tags, correction of tags and number of tags when taggers have much tagging experience and indexing skills. As a librarian, #R and #Jen wondered about numbers of tags for a document. They said "*I cataloged earlier to know that the maximum number of indexed terms is 6 but for tags I could do a little bit more*" (#R).

Secondly, some informants wondered about forms of tags such as abbreviations; tag strings, spelling as followings:

- Adding more explanations for abbreviations: "*LAM is Libraries, Archives and Museum*" (#L).
- Checking spelling of tags by searching right words from Wikipedia, Google, dictionaries and tag clouds. (#L, #R, #C, #K, #A)
- Checking tag string by going back the tag results to see assigned tags (#L).

Thirdly, many informants did not doubt about the correction of tags as well number of tags. The phenomenon happened when:

The informants were confident about their topic experience and they based on the documents to tag. For example, #J was basing on topic experience to tag and had no doubt and consideration about the tag, since "*I don't know. I just tag based on what I have known*" (#R) and "*I will assign tags based on what I know*" (#A).

The informants perceived tagging as free tags. They might have less consideration: "*If I tag, I really don't care how wrong my tags are as I just do anything I like. I do it for myself*" (#R).

Tagging with video and image, most of the informants were lesser doubt about their decision than tagging with text: "*What you see is what you get from the image and video so that I don't worry about tags*" (#Carl). However, for the image or video, if there was no descriptions, taggers didn't know where to check for more information: "*If I want to know the place which appeared within the video, I don't how to do it*" (#Carl).

For tagging with full text documents, most of the informants concentrated on scanning main ideas for tags. They felt confident because they might be sure what the books talk about.

The consideration of tagging decision will be expressed below:

Thought of	Opinion of informants	Behaviors
Number of tags	Many	Looking and checking recall of tags for document.
Spelling of tags	Many	Checking by Google, Wikipedia and dictionaries.
Strings of tags	Few	Looking at the instructions Checking in the tags bar after tagging.
Explanation of tags	Few	Adding explanation Checking the meaning of the abbreviation within text or from the www.

Table 4.16: Consideration of tagging decision

4.4. Type of tags

According to the reviews of types of tags, 8 main types of tags which were studied by Golder and Huberman, (2005); Firan, Nejdl and Paiu (2008); Sterken (2008) through the log analysis and experiments are:

- Topic is what the document is about.
- Time is publishing year for book, days or month of an event.
- Location is places of publishing or places appear in the documents.
- Type of document is fiction, non-fiction or interview.

- Authors/owner is author of books or creator of the video or image.
- Opinions/qualities are tagger's opinion about the book to show up that feeling on the tags such as funny, fantasy or magic.
- Task organization is tagger's way to organize their collection such as: to read or to-do-list.
- Self reference is personal opinion of taggers to take note their documents such as mynote, mystuff or mycomments.

From counting tagging results of nine informants, there are 45 topics and 306 tags totally.

Types of tagging	**Type of tasks**	**Number of tags**
Factual tagging	Article tagging	39
	Full text tagging	68
Fictional tagging	Fictional tagging	65
Entertainment tagging	Videos tagging	66
	Image tagging	68
Total		**306 tags**

Table 4.17: Number of tags

Type of tags	**Topic**	**Location**	**Time**	**Author /people**	**Type of document**	**Personal opinion**	**Task organization**
Factual tags	89	0	0	1	4	0	5
Fictional tags	38	2	1	14	8	10	2
Entertainment tags	86	20	2	5	3	16	1
Total %	69.3%	7.1%	1%	6.6%	4.8%	8.4%	2.7%

Table 4.18: Types of tags

For the factual tags, most of the tags are assigned by topical tags. There are few tags with aim to categorize and organize articles such as: *"science fiction, theme, drama, theory*"; "*Definition, Victoria Government Dpt, Information*". There is no tag of location, time and personal opinion.

For the fictional tags, the numbers of topical tags are still big but the location, time, author/people, and types of document, personal opinion tags are increased. Specially, the personal opinion's tags are increasing in the fictional topics such as "*Suspense, Sad, best-selling fiction, magic, secrecy*".

For the videos and images, the numbers of the types of tags are increased such as topic, location, people, personal opinion and task organization. In the video and image tagging, people usually tagged for places which the pictures and video came from or mentioned within the pictures and video. For example, #C assigned for the picture "*Japan*" where is mentioned within the picture; #E tagged "*Conad Supermarket*" where the pictures has taken.

Compare among factual tags, fictional tags and entertainment tags, the topical tags occupied the most of the types of tags with 69% (213/306tags). While there is no location tag within factual tagging, there are 20 location tags within videos and images tagging. Maybe because of types of document influence on types of tags, the informants took care of the locations of pictures and videos to organize their collections. Moreover, while there is no personal opinion tag within factual tagging, there are many personal opinion tags within fictional tagging (10 tags) and entertainment tagging (16 tags). Maybe because of the academic tagging, the informant focus on topical tags for the academic than personal opinion tags.

From the experiments, the types of tags have some characteristics as follows:

For topical tags, they could be free keywords about any name of projects, diseases, programs, title of books or anything appear in the stories, books, videos and pictures.

- The topical tags may be a general topic from the abstract or the full text. For instance, #A assigned "*Brasov History*" basing on the abstract of the article.
- Comparing to subject headings or controlled keywords, those tags are not at the same of subject levels. They have wide details from very general topics to very detailed topics. For instance, #A labeled for the booklet "Sex safety" with a tag "*Youth sex education*" while #R labeled the booklet with many detailed tags such as

"*Herpes, Genital wants, Chlamydia, Gonorrhoea, Trichomonas vaginalis which are names of sexual diseases*".

- The details of tags are depending on tagging with topic experience, descriptive information and full text. In the table 4.19, tagging with full text, taggers assigned many details tags such as name of projects, details of diseases, names of places than the tagging with abstract and topic experience.

Level of tags	**Tags**
Tags from topic experience	Otaku, videogames, playercreated content, neofiction, user-behaviuor, Kingdoms, monarchies, trilogies, Fantasy, Moses and the ten commandments, The birth of Jesus, The story of Joseph in Egyp, Noah ark, cyberpunk, utopia, dystopia, Hamlet, adaption, drama
Tags from descriptive information	Brasov history, political communication, habermas, deliberation, legitimation, media societies, cosmetics, skin, consumer, protection, products, ingredients
Tags from full text	topic maps, information organization, CEDECA project, cultural heritage, Aquabrowser, (libraries, archives, museums) LAM, HIV, AIDS, STI, Sexually Transmissible Infections, Condoms, Herpes, Genital wants, Chlamydia, Gonorrhoea, Trichomonas vaginalis, Youth sex education, Safe sex, Sexual transmissible infection (STIs), Sexual health, Sex protection, Contraceptives, Condoms. transmissable diseases, clinics, STI clinics, symptoms, treatments, risks, information strans dis, oral sex, help, guide, help.

Table 4.19: Topical tags from topic experience, information descriptive and full text

For the tags about author or people mentioned in fiction, the informants assigned very few authors tags for factual topics while they assigned many authors tags for fictions and many people which appeared in the fictions such as "*dumbledore, voldemort, ron weasley, hermioni*". Additionally, comparing to videos, images have more people tags as the informants tagged for people appeared in the pictures or they tags for whom they want to

share with such as "*Ezerea, Andrew, Yalda*". Specially, the name of person in the text or videos and images are assigned dissimilarly. Instead of Dumbledore, #E tagged "***Professor*** *Dumble Doe*"; or #E tagged for some her friend's names such as "*Andrew, Yalda*" which she didn't assign full names and formal names.

For the time tags, comparing to time in classical taxonomy, taggers performed time in different ways. While #J assigned 1997, #W assigned "Pre-colonization" or #E with "2010 songs". Moreover, the informant also added text into the time such as "*Miley Cyrus 2010 songs*".

For the locations tags, the methods to display the name of places are heterogeneous.

- Different names of a place such as "Nordkap" instead of "North Cape; or "Old town" with aim to "Tallinn".
- Personal ways to name for places such as "Northern-Most-Part-Of-Europe" which pointed to "North Cape" or "countryside", a place in the Ireland.
- Variety of details of location names. It could be a general name of a continent which indicates to a specific place such as "Latin American" point to "Colombia"; or "Asia" point to "Japan"; or "Greenland" point to "Denmark"; or "Conad supermarket" point to "Parma".

For the personal opinion tags, the informants could put their feelings or their evaluation on the content of documents. The informant usually put the personal opinion tags for videos, pictures and fiction rather than academic documents. For the evaluation tags in the table 4.19**,** the tags might be useful to recommend the qualities and interests of documents for other seekers like "*famous music, famous dancer, Colombia in passion, love/inspirational song*".

Evaluation tags (quality tags)	**Feeling tags**
Colombia is passion, Love/inspirational song, Troubled teenager, Young love, Childhood innocence, famous music, popularity, famous dancer	Idiot, Optimistic, funny, Uplifting, Cold, Scenic, big, fun, Suspense, Sad

Table 4.20: Personal opinion tags

For the task organization, the informants could assign tags with aim to organize their collections. The task organization might be a name of program, website such *as "Kaffe med Kunrt (Coffee with Kurt), Forvalsadvokaterne.com*" or their personal reference such as *"To-do-list*" or notice about documents such as "*Volume, definition, information*".

4.5 Factors influence on tagging behaviors

According to Sen, Lam, Rashid, Cosley, Frankowski, Osterhouse, Harper, and Riedl, (2006), there are two major factors impact on tagging behaviors: personal tendency and communities' tags. Besides, Rashmisinha (2005) has pointed out that there are different tagging decisions making in various cultural knowledge. Additionally, Bar-Ilan, Zhitomirsky-Geffet, Miller and Shoham (2010) concluded the exploration of metadata of taggers will affect to result of tags. Furthermore, my study also figured out some other factors such as indexing skill, tagging experience, tagging objects (image or video or text) and some opposite opinions in the community's tags.

4.5.1 Personal tendency

Sen, Lam, Rashid, Cosley, Frankowski, Osterhouse, Harper, and Riedl (2006) defined that personal tendency is their preferences, knowledge, experience and beliefs about the tags. According to my study, the personal tendency is one of the most important factors strongly influenced on tagging behaviors such as explorations, motivations and tag formulations. As the result, most of the informants were powered by their knowledge and beliefs about the tags in the experiment: "*I really don't care how other people think of my tags so I just tag base on what I've known and my feeling*" (#R).

4.5.2 Communities' tags

Sen et al. (2006) found that the community' tags strongly persuades taggers assign tags through the compare of experimental tags in Movielens. The influences are performed by user's vocabulary and user's first tags. Nonetheless, this statement isn't fit with the study because of some reasons as follows:

- While the above authors used tags in the Movielens to compare each others, my study observed the whole process of tagging behaviors either in text document tagging or video or image tagging.
- Besides, the authors just observed the tagging outcomes to conclude the relationships between personal tags and community's tags. Thus, there is a concern about whether the tagging outcomes are representative for the relationships of personal tags and community's tags. For example, during tagging in CiteULike #R saw the suggested tags *"user-behavior*" but he had never looked at it for his tags. However, in this tags later, there was a tag "*user-behaviour*". In this case, it is impossible to say there was relationship between the suggested tags with his tags.
- The fact that, there were few informants who were attractive by community's tags. When #L, #R and #W looked at the suggested tags, they didn't choose the suggested tags for their tags: "*I don't think these tags are good for me*" (#Cark); "*Ah ha, they have tags already, but I want to tag by myself*" (#R).
- Few informants used the suggest tags for their own tags. For example, when #L didn't know how to tag for book made into film, she came to the tag clouds in LibraryThing to follow other taggers.

In conclusion, when taggers feel confident to understand about topics and documents, they rarely are affected by the community's tags. On the other hand, taggers might be affected by the community's tag when they don't know what they should tag for documents which they are not sure of the contents.

4.5.3 Indexing skills

Librarians who have knowledge and skills of indexing, information seeking as well as folksonomies also influenced on the tagging behaviors. Some librarian informants were clear about how to understand a topic, numbers of tags or assigning right terms which are searchable to others: "*Because I am a librarian, I won't give so many tags because of spreading of documents*" (#R); or searching well-known terms in the LibraryThing's tag cloud for her tags as she said "*Doing it, other people can search on my tags easily*" (#L).

However, this factor does not mean that if taggers are not librarians, they won't have sophisticated tagging behaviors:

- #C, who is a non librarian, realized that if she wants to archive only, she doesn't need to have synonyms for the tag "Kids": "*I am wondering that these tags can be searchable by other people so that I need to have "Kids" and "children. If I just want to keep this book in my library, it is no need to have children tag. But okay, I can keep it"* (#C).
- #P is a librarian and none tagging experience wasn't clear about tagging motivation, direction of how to assign a tag, how to have good tags for organizations or sharing. For example, she tagged for most of the documents with such kind of tags: *"This-book-need-to-go-back*" or *"This-book-is-interesting-to-read*", etc.

4.5.4 Tagging experience

The tagging experience has influences on tagger's exploration, decision making and tag formulations. For example, most of the informants, who claimed from excellent to fair tagging experience, knew how to tag, motivations for tagging and they met less difficulties in tagging. On the other hand, #W who claimed with none tagging experience wanted to give up tagging in tagging in LibraryThing and CiteULike. She said "*I don't know what should I tag for the article" or "I am a new tagger so that I don't know how to tag in good way*".

4.5.5 Tagging objects

According to my study, most of the informants feel more comfortable to tag with videos and images than when they do with text tagging. For video and image, most of the informant met less difficulty in understanding, decision making and tag formulation. In the interviews those informants said video and image are easier to understand as what we see is what we get for tags. Moreover, for the text, #Carl in the pilot study said "*I am lazy to read the text compared with watching an interesting movie or a picture*". This point doesn't mean the video and image is definitely easier than text document. This means that people who get acquainted with types of documents might feel familiar to tag within those specific types.

Moreover, there are various types of tags through different types of documents. For example, for video and image, the number of personal tags and location tags are increasing while there is no personal and location tags within articles and full text tagging.

4.5.6 Cultural knowledge

The cultural knowledge also affect on the decision making: "*Categorization is often based on cultural knowledge. For example, over the years we learn the cultural consensus regarding the boundary between wolf and dog, couch and chair, fruit and vegetable*" (Rashmisinha, 2005, p.3). However, she argued that in the digital environment, the factor is less influenced on the tagging decisions as taggers focus on objects such as books, articles or movie, images.

It is difficult to know this factor in tagging because it relates to cognition and personal background. Thus, this factor will be discussed under two main types of taggers: confident taggers and certainty taggers. The first taggers group assigns tags basing on topic experience whether they read documents or not. The second taggers group assigns tags basing on either topic experience or document exploration.

4.5.7 Exploration of detailed metadata

Bar-Ilan, Zhitomirsky-Geffet, Miller and Shoham (2010) found that there are different tagging results among a variety of exploration of detailed metadata. The detailed metadata means to the level of information which is read by taggers. They have found that the more details of pictures people read, the more detailed tags were given. Consequently, the authors concluded that the more interaction of websites which include the pictures, the more tag numbers and detailed tags.

In the experiment, the confident taggers who didn't seek any information have lesser number of tags than the certainty tagger who sought information to understand. Moreover, there are different detailed tags between of tags based on topic experience, descriptive information and full text. For example, basing on full text reading, #L got many detail of project names in the article; or #K said "*Compared to full text tagging, I will tag with*

abstract differently". Explaining about this behavior, #K said "*Because the abstract might have less information than the full text*".

4.6. Tagging challenges

According to the analysis about tagging behaviors process above, there will some tagging challenges such as: information resources issues, tagging functions issues, vocabularies supporting issues and patience in tagging issues.

4.6.1 Information resources issues

4.6.1.1 Unattractive information structure

Through the explorations stage in LibraryThing, there is prominent problem that the information structure in LibraryThing is not helpful for informants in reading and exploring and understanding. On the other hand, most of informants agreed that the simple tagging site as CiteULike made them easier and more comfortable to tag than LibraryThing. Even though LibraryThing has many useful functions which help users understand about the books, 8 of 9 informants were not attractive to read in LibraryThing. Only #C came to read the descriptions of fiction.

Figure 4.21: LibraryThing tagging page

Most of the informants have a common behavior in the stage of explorations including reading abstract firstly. However, LibraryThing has descriptions at the bottom of page or in the small place in the page where people were hard to see immediately, while CiteULike has abstracts directly under the title of article.

4.6.1.2 Poor metadata

Since the metadata are the main elements for taggers to look before tagging decisions such as title, abstract, author, year publishing and so on, it affects strongly to the certainty taggers who mainly based on reading and seeking from the metadata. It made #A in the Brasov history topic, #W in the Romeo and Juliet drama topic feel confused and want to gave up exploration as the abstracts were less and unclear information.

In the YouTube, there is less information about place of video, time of video, topic about videos which are up loaded by users while Flickr have better information for tagger to look such as short descriptions about the pictures and places of pictures. As a result, when #J would like to know the genre of the song, he was disappointed that YouTube didn't have it.

4.6.2 Tagging function issues

4.6.2.1 Quick tagging functions

For confident taggers, they need a quick function to tag instead of going step by step from the explorations to tagging decisions. While CiteULike has offered such kind of function, LibraryThing requires tagger go through many steps. Moreover, most of informants said the tagging function in LibraryThing is wordy and time consuming. In the figure 4.22, 4.23 and 4.24, we will see an illustration of quick tagging function and normal tagging functions in CiteULike and LibraryThing:

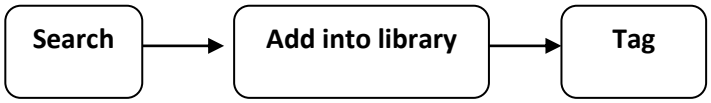

Figure 4.22: Quick tagging steps in CiteULike

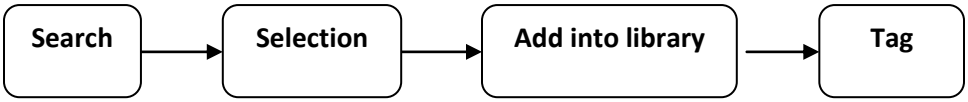

Figure 4.23: Normal tagging steps in CiteULike

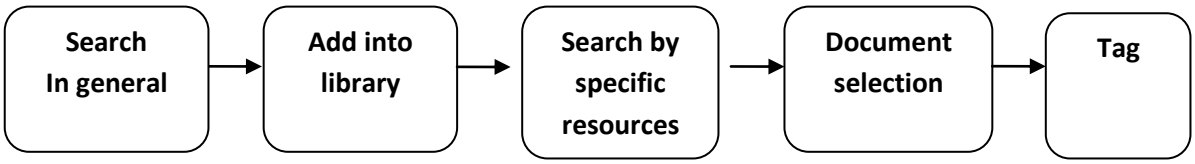

Figure 4.24: Normal tagging steps in LibraryThing

4.6.2.2 Tag strings combinations

During assigning tags in CiteULike, most of the informants struggled with tag strings combination. In CiteULike, there is a rule that one tag which has more than one word should have hyphens to combine them together. "*this-is-funny-and-exciting-book*" is an example. The CiteULike also offered short instructions about the rule but nowadays, they have removed it. Nevertheless, most of the informants didn't pay attentions to the instructions, except #L. #L read it but she still felt unclear in the strings combination. She said "*the instruction is not good"* and *"They should notice about the tag combination*". On the other hand, there was no problem in tag strings combination in LibraryThing as they allow informants to assign tags in natural language order.

4.6.2.3 Poor suggested tags

Some taggers wanted to look at the suggested tags from the communities or tagging system but many articles in CiteULike have never been tagged. While LibraryThing has offered either personal tag clouds or entire tag clouds of the library, CiteULike just only offered personal tag clouds. The entire library tag cloud helped #L to choose tags for her book in LibraryThing.

4.6.3 Vocabulary issues

4.6.3.1 Misspellings

According to semantic of tags, there are 9% of miss-spelt tags. For example, instead of "*transmissible diseases*", #K has written "*transmissable diseases*". The miss-spellings have follows characteristics:

- Normal misspelling such as instead of "*transmissible*", #R wrote *"transmissable".*
- Wrong string combinations. For example, #R wrote "playercreated" instead of "Player created".
- Personal abbreviation. For example, #K wrote "strans dis" instead of "transmissible diseases".
- Local languages. For example, #R used some Danish tags such as Iniut or Hoeg.
- Wrong names such as J. Rowling, Voldemort, and so on.
- Many informants have wrong use of capital letter. For example, with some names, they still wrote in lower case written.
- Undefined mistakes. For example, #R wrote "neofiction" but he said that he didn't know what he means to the tag. This makes the tag become more ambiguous.

In case there was wrong spelling notice, sometime he came to dictionaries to check but sometime he didn't want to it. He said that he noticed about the red color under miss spellings but no such kind of systems can suggest immediately right words for him as Microsoft Words does now.

4.6.3.2 Difficulties in term understand

By observations of experiments, during reading abstracts, full text or suggested tags, many informants wanted to search for meanings of the terms which appeared in the text. For example, #R came to Wikipedia to check meaning of "*Trichomonas vaginalis*" or #K came to Google to check meaning of STIs term.

In talk aloud, #R said it is better if tagging sites offered functions to support them in case they need to understand terms, by clicking on that term, there will be explanations

appearing. This idea leads us to semantic web where system can have more links to each important term in the text. Wikipedia is a case in point. Moreover, in the pilot study with #Carl, she said she would be clear and feel convenient with explanations which have visible examples and pictures to see.

4.6.3.3 Difficulties in tags extractions

During reading abstract or full text, almost all informants have the same common behaviors that if seeing some interesting terms or right terms with their ideas, they wanted to copy them into their tags. However, there is no automatic function with aim to support for this. Thus the taggers need to do it manually in step by step. As the result, the complicated step to copy terms into tags might make taggers loose attention to tag formulation.

In the talk aloud, #R suggested that it is better if the systems can allow tagger to extract automatically into their tags when they click or double click on the terms which appear within text. It is not only support taggers decide tags quickly but also help them focus on reading to understand documents.

4.6.3.4 Ambiguous tags

Spitieri (2007) used document analysis to conclude that the tags are ambiguous and inconsistency because they are misspelling mistakes which illustrated in the miss-spellings analysis above. Moreover, they found the tags might be abbreviations, initialism and acronyms. For example, some taggers may abbreviate SF is Science Fiction but others may refer it to San Francisco. Besides, the tags are also neologisms, slangs and jargons which bring both benefit and drawback for users and systems. On one hand, it could make good classical taxonomy's shortcomings in regards of up-dated tags and friendly tags. On the other hand, they are also ambiguous and inconsistency as they are not unique and verified commonly. In the experiments, the informants also met those issues.

Firstly, sometime, taggers didn't know what they mean to tags. For example, #R has tagged for the videogames with "neofiction" but in the interview, he said "*I don't know what that*

mean". Checking the descriptive information of his selected article *"Playing with videogames*", there was no "neo-fiction" within the text.

Secondly, many informants met difficulties in selecting right tags because they might consider about synonyms. For example, #C considered to tag and kept or deleted one of the tags: "*children*" or "*kids*". Furthermore, Spitieri, L. (2007), Golder and Huberman (2005) showed that the ambiguous tags are not only come from synonyms but also so come from polysemy and basic variants. For example, polysemy may refer to a hole in the wall or to a pane of glass.

For the ambiguous abbreviations, from the survey of LibraryThing, they offer solutions for system can recognize the association among them (NF might be non-fiction) and using combination and voting for doubt tags (SF might be science fiction or San Francisco). On the hand, CiteULike doesn't have such kind of solutions.

For synonyms, polysemy or variants of tags issues, Ames and Naaman (2007) said that the tagging system should not force tagger to tag basing dictionaries or any library's thesaurus. They argued that doing that, taggers could be influenced by classical taxonomy systems while the positive point of tag is avoiding the old system's drawbacks. On the other hand, Spitieri (2007) suggested 2 possible solutions as follows:

- A link to recognize online dictionary or thesauri as well as Wikipedia for user's decision of tagging. In CiteULike, LibraryThing, Flickr and YouTube, there no such kind of functions which have been integrated by Wikipedia or Google or thesauri or dictionaries, while most of informants need these functions.
- The cataloging system creates clear- written recommendations for choice and form of tags which include different types of semantic and components tags. At this point, LibraryThing could be such kind of this example, as they use communities' tags to recommend and use related subjects from traditional catalog to suggest tags for taggers.

In conclusion of these challenges, according to #L and #R, they suggested that it is better if the tagging sites support them by thesaurus or control vocabularies. It means the systems

could integrate or have specific solutions for vocabulary control as a reference for taggers. However, Ames and his colleague (2007) and Spitieri (2007) wondered that the ways to combine and make use advantages of either folksonomies or taxonomies are a challenge for tagging nowadays.

4.6.*4 Patience issues*

According to the experiments, patience is an important factor influence on tagging results. It is clear that there are a variety of patient capacity in 9 informants through explorations, tag formulations and tag checking:

- One informant tried to find good tags and she wanted to check her tags after tagging.
- There are many informants tried to explore for understanding about documents. However, if there were trouble or difficulties, they gave up exploring. Interviewing these people, they said it is better if the system have simple and effective function for them to read, understand and tag for documents. For example, it should have simple information structure as CiteULike and should have more visible for text tagging as well as easier tagging function as CiteULike.

In the stage of tag formulations, most of informants are patient to formulate their tags while very few informants want to check the tags after tagging, except to #L. This behavior maybe leads to ambiguous and inconsistent tags.

Conclusions:

From the entire analysis above, the study about process of tagging behaviors has found some major findings:

1. Tagging experience and perceptions:

Most of the informants are aware of tagging such as what is tag and how to tag it. However, there is a variety of tagging experiences. For example, they may be experts in tagging study. They may be practitioners in tagging sites such as Facebook, YouTube, Flickr, Picasa, etc. They may be a beginner who has a little knowledge of tagging but they never practice frequently. Those factors have strong influences on their process of tagging.

2. Classification of taggers:

There are 2 types of taggers: confident taggers and certainty taggers. The confident taggers intend to tag basing topic experience no matter what whether they read the documents or not. The certainty taggers intend to tag basing topic experience and document exploration. The classifications will lead to 2 different processes of tagging behaviors.

3. Explorations:

Investigating among 9 informants, there are 3 main aspects in the explorations: confident seekers, certainty seekers and exploration of video and image. For the exploration of confident taggers, they usually browse the information but they never read carefully because they memorized what they have read before. For the exploration of certainty taggers, they tried to look metadata and check in the World Wide Web to understand of documents. For videos and images, the explorations are simpler than the text tagging. They just watched and they understood easily.

4. Tagging decisions:

Following the explorations, there are 2 main formulations of tagging decisions: decisions of confident taggers and decisions of certainty taggers. Firstly, the confident taggers base on topic experience to decide tags while the certainty taggers base on both topic experience and document exploration to make tagging decision. Secondly, there are the connections between tags and searching phrases: accurate match and hierarchy relationships. Thirdly, there are 3 types of tagging motivations such as self organization, social organization and self communication. Fourthly, while the certainty taggers have much consideration to decide tags, the confident taggers feel comfortable with their tags as they base on their memory of topic to tag. Fifthly, taggers feel simpler and comfortable to tag with video and image than the text, because what they see from them is what they get to tag.

5. Type of tags:

The analysis found that most of the tags are topical tags. Firstly, while fictional topics and entertainment topics have diverse types of tags such as topical tags, authors, years, types of

documents, location, task organization, and more personal opinions tags, the factual tags have mainly topical tags and very few tags of task organizations and types of document. Secondly, the tags are diverse among taggers. For example, with the same topic but tags are assigned differently in various taggers. Thirdly, it is different ways to recall name or time or people such as "*2010 song*" or "*Post-colony*".

6. Factors influence on tagging behaviors:

Combination between tagging theories and the results, there are several factors influence on tagging behaviors such as: personal tendency (tagging experience, indexing knowledge, topic experience and motivations); community's tag which is not so strong influences on tagging behaviors; tags recommendation algorithms; tagging objects such as video or text; various cultural knowledge and exploration of detailed metadata.

7. Tagging challenges:

There are 4 main challenges of tagging. The first is information resources issues such as unattractive information structure, poor information descriptions. The second is tagging function issues such as quick tagging solutions, tag strings combination and poor suggested tags. The third is vocabularies issues such as miss spellings, terms issues, extractions; ambiguous tags. The final is patience in tagging.

CHAPTER 5: CONCLUSIONS AND RECOMMENDATIONS

The chapter presents the usage of findings of the study to answer for research questions. Then, some recommendations and research limitations as well as further researches are also discussed.

5.1 Discussions of the research questions

The major aim of this study is investigation of process of tagging behaviors and challenges in tagging behaviors. To achieve these purposes, two research questions have been formulated at the beginning of the study.

Research question 1: *What are user behaviors characteristics in tagging process?*

Research question 1: *Which challenges do the taggers meet during tagging processes?*

5.1.1 Discussion of research question 1

Throughout the chapter 4, there are two main types of taggers, confident taggers and certainty taggers. **The confident taggers** are the informants who assign tags basing on only on topic experience. **The certainty taggers** are the informants who assign tags basing on both topic experience or document exploration. Additionally, tagging with video and image differs to tagging with text documents.

There are several factors influence on tagging behaviors such as personal tendency, cultural knowledge, and exploration of detailed metadata, topic experience, indexing skills, tagging experience, community's tags and tagging objects.

5.1.1.1 Tagging behavior process of the confident taggers

The tagging behaviors of the confident taggers have two stages: scanning and memorization; and tagging decision.

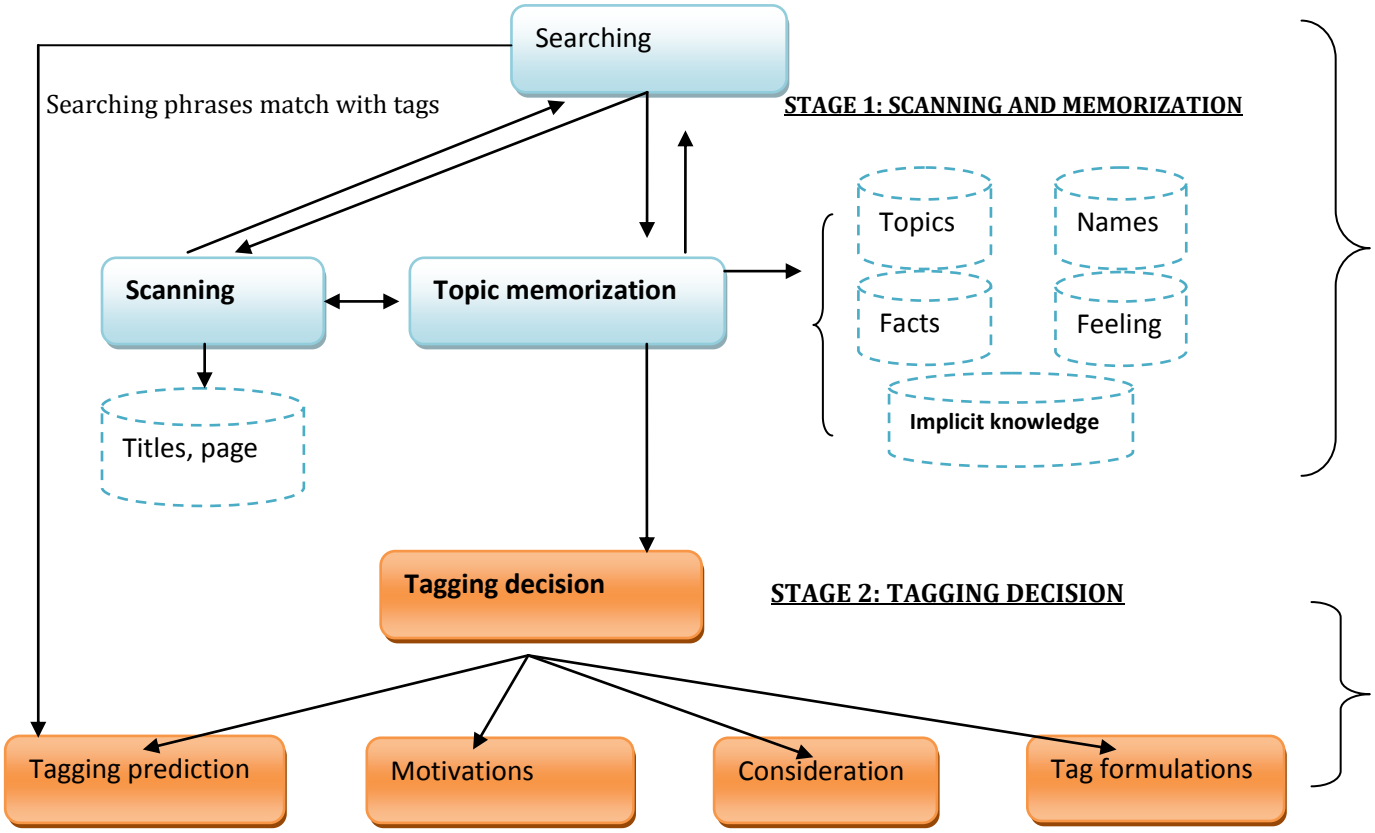

Table 5.1: Tagging behavior process of the confident taggers

In the first stage, the confident taggers tried to memorize about topic of the documents after scanning documents. In some cases, taggers might return to the searching stage to refine suitable documents. The memorization is the most important step for the confident taggers to decide tags. Consequently, the taggers might be able to remember topics, facts, names, feeling of the documents. Moreover, there might have some implicit knowledge about topic but they could not recall in the first stage. The implicit knowledge might be recalled in stage of tagging decision.

In the second stage, the confident taggers had quick tagging decision which was based on the topic experience after quick scanning and memorization. Most of the confident taggers had their predictions through the matches between tags and searching phrases. They had motivation to organize documents for archives and searching findability. For the tag

formulation, the tags were extracted from the tagger's languages and topic memories. As the result, the tags were strongly affected by taggers' topic experience. Furthermore, the confident taggers also considered about correction of tags, number of tags but the consideration would not be verified by any information seeking.

5.1.1.2 Tagging behavior process of the certainty taggers

For text tagging, there are three stages of tagging behavior: searching and memorizations, exploration, tagging decision.

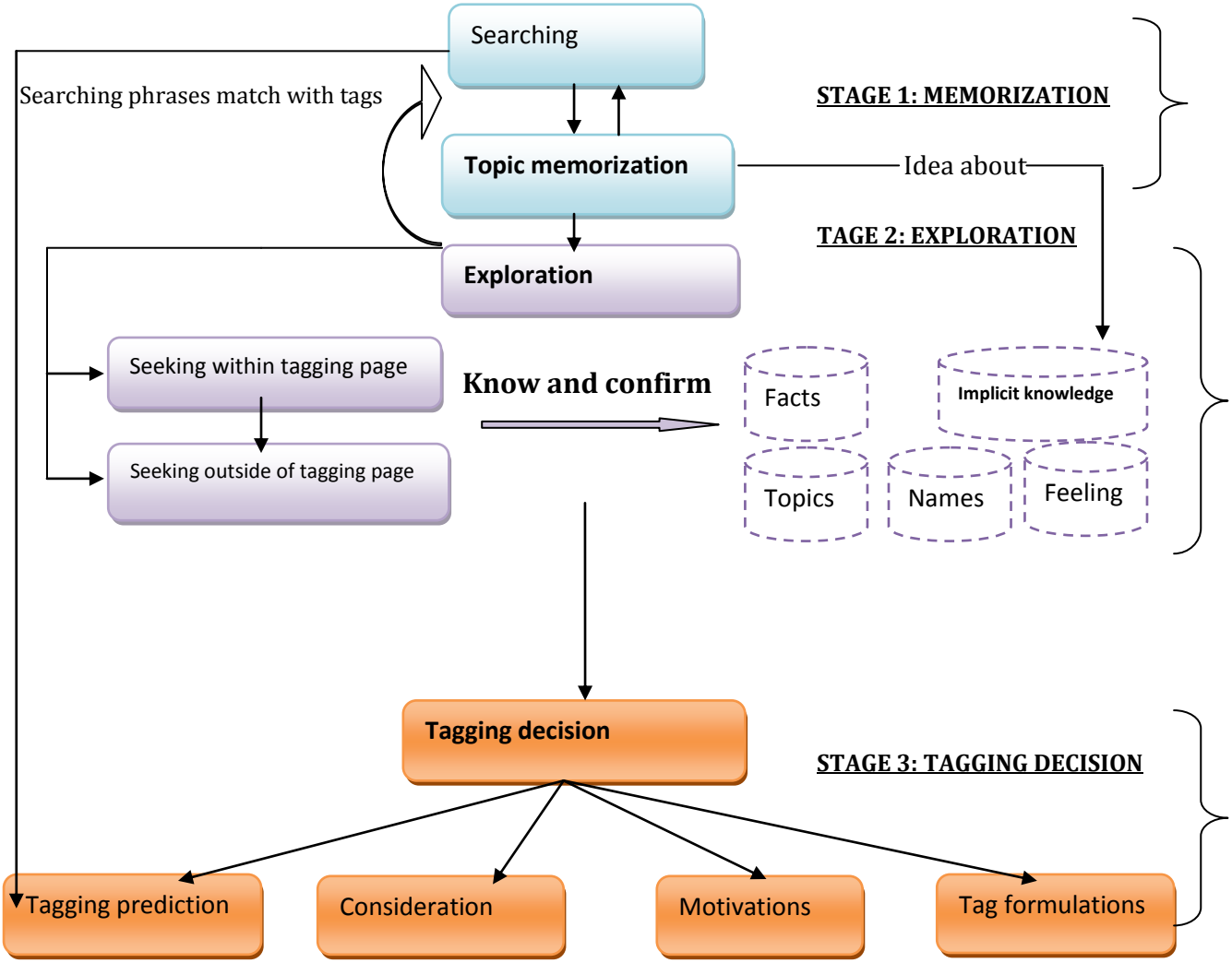

Table 5.2: Tagging behavior process of the certainty taggers

The first stage was topic memorization. At this stage, the taggers recognized document's topics and memorize what they knew about it. In some cases, taggers might return to the searching stage to choose other suitable documents for tagging. At the same time, they were narrowing or deducting the document's topic to their memories. Apparently, they had clearer directions of methods to read and explore in the exploration stage. The taggers might have idea about topics, names, fact and feeling as well as implicit knowledge like the confident taggers.

The second stage is exploration. As the result of the first stage, the certainty taggers will read and understand more about topics, names, fact, feeling and some implicit knowledge before tagging. There are two kinds of exploration, exploration inside of tagging systems and explorations outside of tagging systems. Titles and abstracts are two main things to be explored in the tagging systems. Definitions, information about the book, name of people, place as well as linguistic issues will be explored through Wikipedia, Google, Google book and other local dictionaries. The observation shows that many taggers preferred searching and checking in Wikipedia, Google or dictionaries as they are simple, effective and satisfied with their information demands than the tagging systems.

The third stage is tagging decision. After exploration and understanding of documents, the certainty taggers decide to tag in confident feeling since they know of topic clearly. The tagging decisions based on both topic experience and document exploration. Most of the certainty taggers have predictions through the matches between tags and searching phrases. They motivate to tag for social organization through using synonyms, explanations of abbreviations and correction of tags. This tagger group has much consideration during tagging such as formats of tag, number of tags, correction of tag and the uses of tag. To do the corrections, they try to read abstracts, full text, tag clouds as well as any metadata inside and outside tagging pages.

In conclusion for text tagging, the process of tagging behavior will be modeled as follow:

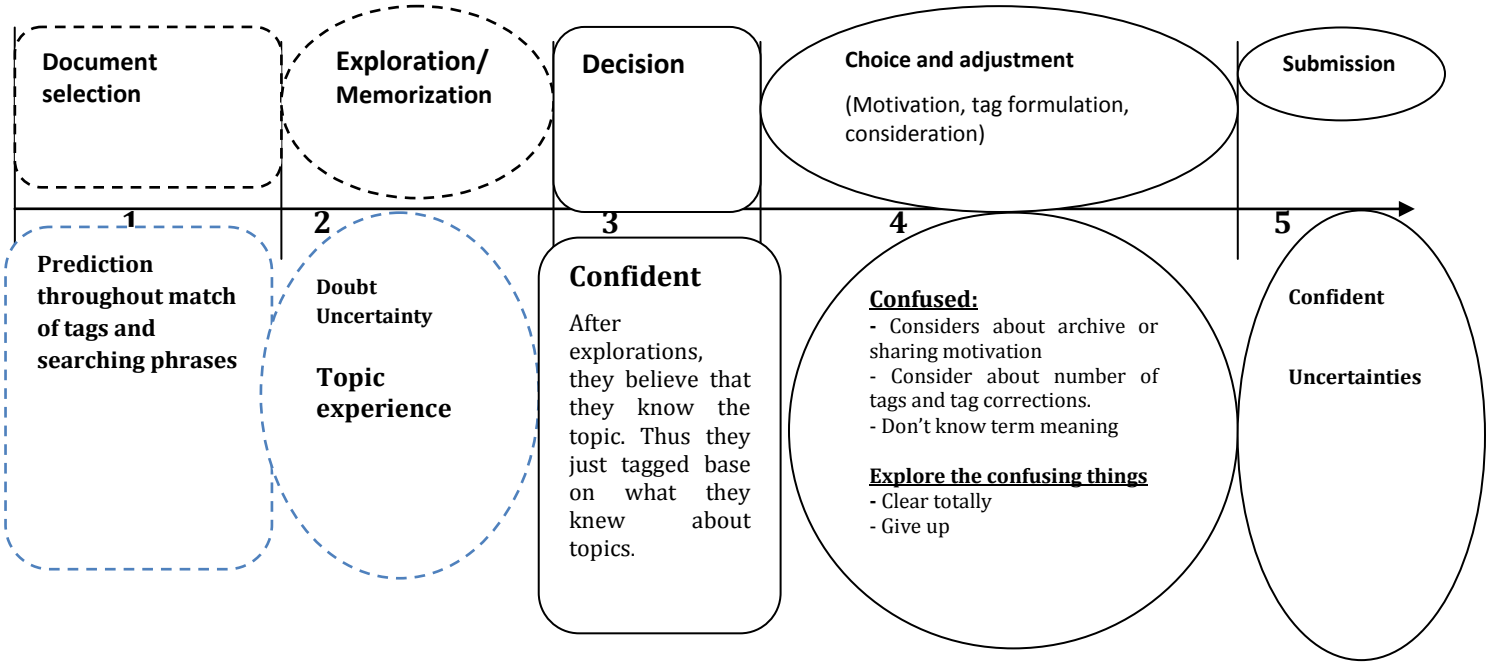

Table 5.3: Text tagging behavior process

The certainty taggers go through the entire steps, while the confident taggers spend on stage of 1, 3, 4 and 5. Firstly, in the document selections, taggers may have tagging prediction through the matches between tags and searching phrases. Secondly, the doubt and uncertainty come out in the exploration or memorization step. At this time, the certainty taggers might try to look titles, abstracts, a scan of full text, tag clouds as well as other information from the World Wide Web, while the confident taggers memorize topics they have known about documents. Thirdly, taggers decide to tag basing on what they understand of documents. At this stage, taggers might feel confident as they thought they would tags basing on what they get from the exploration step. Fourthly, taggers adjust choices of tags: tagging motivations, topics, names and fact, etc. At the mean time, taggers might feel confused and doubted again on their choices. Thus, some taggers will try to correct the confusing and doubt while other taggers might give up doing that. Finally, taggers decide to submit their tags. Some of them may feel comfortable and confident to submit tags as they explore and correct the doubt or consideration. On the hand, the others may submit tags in uncertainty situations because the taggers give up exploring the doubt and consideration.

5.1.1.3 Tagging behavior process of video and image

Comparing to text tagging, video and image tagging are simpler and more comfortable. It has several steps such as topic experience, watching and memorizations, tagging decision which includes tagging motivation, tag formulation and tagging consideration. Firstly, the watching and topic memorizations appear simultaneously. The video and picture watching help them memorize easily by interesting images and sounds. Moreover, they also try to narrow the topic into their experience and knowledge about it to tag. Secondly, the tagging decision comes up their mind during watching and they type down tags as the same time. To explain this behavior, many participants say that "*what you see is what you get*" from the video and image since they are visible and enjoyable.

According to the prominent characteristics of video and image, their tags base on things appeared within the videos and pictures. Some taggers want to know about name of places, people, things in video and image. They may explore them in the tagging systems in YouTube or Flickr as well as go to Google or Wikipedia to understand more about name of place, people, music genre, etc.

In conclusion for video and image tagging, the process of tagging behavior will be modeled: as follow:

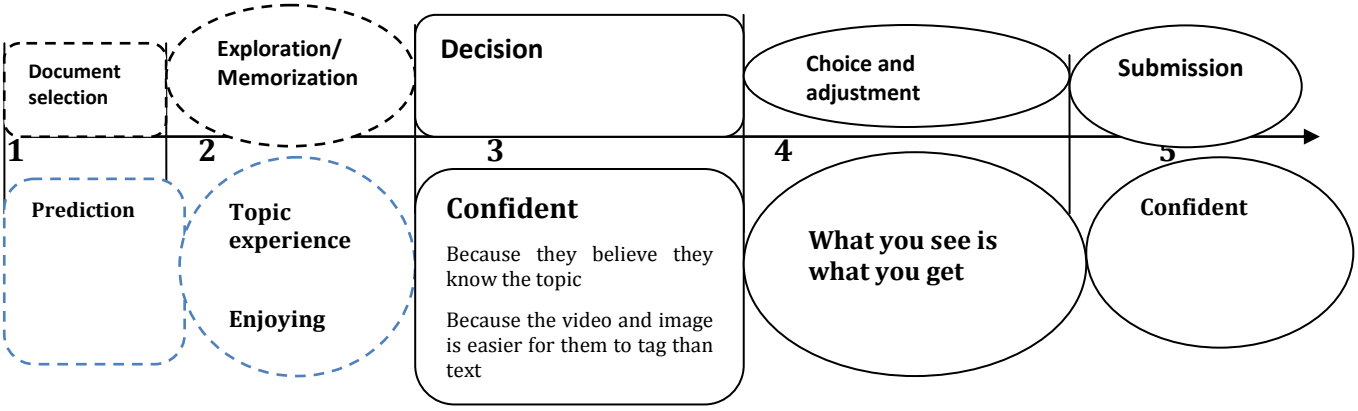

Table 5.4: Video and image tagging behavior process

5.1.1.4 Compare of the text and video, image tagging processes

Comparing between the table 5.3 and 5.4, the text tagging process have similarities and dissimilarities with the video and image tagging process. The stage of document selection and tagging submission are similar for both types of tagging processes. On the other hand, those tagging processes are different in the stages of exploration and decision making.

Firstly, the tagging prediction and the tagging submission have the same characteristics in both text tagging and video, image tagging. The analysis showed that most of tags in text and video, image are matched with the searching phrases. Moreover, in the submission stage, the informants might feel confident or uncertainty about their tags. Most of the informants didn't check the tags after submission. Only #L came to her library to check the tags after the submission.

Secondly, it is clear that the exploration and tagging decision are different in text tagging and video, image tagging. In the exploration, most of the informants watched contents in the videos and images to understand with less doubt and consideration, whereas the informants needed to seeking, read text documents with much doubt and consideration. In the tagging decision stage, most of the informants felt easy to choose terms, formulate tags and less consideration about tags with the video and image. On the other hand, most of the informants were confused about tagging motivation, tag formulation and tagging consideration about correction of tags, numbers of tags and format of tags in text tagging.

5.1.2 Discussion of research question 2

There are 4 tagging challenges. Firstly, there are issues in Information resources such as unattractive information structure in LibraryThing and Poor information descriptions (bad abstract quality and less descriptive information from YouTube and Flickr). The second is tagging function issues. For instance, quick tagging functions in LibraryThing has not offered for tagger type 1; tag strings combinations in CiteULike made tagger confused and ambiguous tags; and poor suggested tags from CiteULike. The third is vocabulary issues such as miss-spellings, difficulties in term understand and ambiguous tags. Finally, patience

issues also effect to tagging behaviors and tag quality. All the challenges are analyzed in detail in the Chapter 4, part 4.6.

5.2 Recommendations

5.2.1 Improving information resources

Learning from CiteULike, LibraryThing, Flickr and YouTube, there are two criteria to have good information structures for tagging: simple and effective. To achieve these criteria, the tagging system could offer this kind of information:

- Good abstracts and reference links such as dictionaries, Wikipedia or full text.
- Having more suggested tags as LibraryThing has done.
- For video and image, it is better to have more information about place, name of people appear in the video and picture.
- Having pictures for text documents which can be representative for the documents: "*It is good if CiteULike or LibraryThing can be included some representative pictures for the books. For example, when mentioning to Harry Potter, people easily recognize a Harry Potter, Wizards…*" (#Carl).

5.2.2 Improving tagging functions

Learning from the CiteULike, tagging system should have quick and flexible tagging function which support taggers to tag documents by many levels: at the beginning stage or after explorations stage.

Avoiding from the CiteULike, tagging system should have clear instructions about the way to give tags. This instruction could appear in the tagging page which could help taggers easily to get it. Besides, the site could display the instructions by text or pictures or video. According to many informants, they are more interested and easily understandable with pictures or video than text.

Furthermore, learning form Librarything, system should allow taggers tag in natural written way as it is friendly than putting tag strings in combination rules as CiteULike.

5.2.3 Improving tagging support

According to the vocabulary issues, it is better to have such kinds of functions to support tagger tag in convenient way:

Firstly, there should have miss-spelling notice machine to alert tagger about wrong spellings. From that, after miss-spellings notice, system can suggest right terms or synonyms as Microsoft Words has done.

The second is a solution of term extraction automatically which supports taggers to register, extract and tag when tagger clicks on terms from the text. Chi and Mytkowicz (2007) suggested a model to support vocabulary issues in tagging. The idea is creating a function in which user can click on words of paragraphs to keep them as tags later. There are some other solutions to avoid vocabulary issues such as Lazy Sheep's model which auto-tag and auto-describes user's bookmark. However, these authors argued that it could lead taggers into cognitive barriers of tagging. To solve it, user communities' tags could be good solution to adjust this argument.

Thirdly, tagging should have the integration of annotations for academics, unfriendly terms and important terms which appear in the text into the system. For example, if taggers don't know about a specific term in a specific subject, there will be an explanation appears when taggers click on it. This idea has the same Wikipedia solution of semantic web to connect term with definition:

> *"If they could highlight some important terms which are popular or representative for the documents in the abstracts or anywhere, I will easily to be attractive and memorable*" (#L)

Finally, Yin, Xue, Hong and Davison (2010) based on Bayesian approach to predict user's tags through user interest or seeking behaviors. Through the experiment, the model can improve the system to suggest tags to right person, right content. They also noticed that the tagging system weights on user interests as well user profile. However, it still has issue that do users give enough personal information for system to do this work.

5.2.4 Improving tag ambiguities

For ambiguous abbreviations, learning from LibraryThing, tagging system could be able to recognize the associations among them. For example, they can put SF and Science Fiction closes together. Besides, the system may let users vote for: SF is Science Fiction or San Francisco.

To avoid the ambiguous tags, tagging system should notice it when tagger assigns tags at the beginning. For example, when they type SF, there will be some options for meaning of SF appeared to choose. Doing it, there are two approaches, one base on system and one base on user created content. However, to have user contributions, the tagging system should build such kind of solutions.

For synonyms or basic variants of terms, tagging system may solve it by thesauri links or integrate into tagging system,

5.3. Research limitations

According to data collection and data analysis, there are three major research limitations as follows:

Firstly, the experiment cannot observe from real taggers in real tagging environments because data might be not completely validated. The real taggers will have clear motivations on tagging so that they know what they should tag and how to tag. On the hand, informants may do tagging in an inconvenient way since they don't know how to tag and why they should tag. For example, most of the informants are real taggers within video and image tagging, they felt easier to tag with the videos and pictures rather than text tagging.

Secondly, tagging experiences on experimental tagging sites is an issue. According to the samples' tagging experience, none of them have experience to tag in CiteULike and LibraryThing. As the result, most of them were confused to read, search and tag in those sites. For example, #A said she has never done with CiteULike or LibraryThing so that she doesn't know how to tag.

Last but not least, there is linguistic limitation, because most of the informants are not native English speaking people. The experiments were conducted in English which expected the informants to finish all activities in English such as search, read, understand and tag. Especially, talking aloud by English, they may not express all the idea and feeling by English. Moreover, as they are not English speaking people, their voices or pronunciations will make difficulties in transcriptions.

5.4 Further studies

Firstly, as an experimental study about tagging, the author cannot observe real taggers in a real tagging environment. Thus other researchers might use interviews or semi experiments to investigate the process of tagging behaviors within narrow tagging system, broad tagging system, text tagging, video, image tagging, academic tagging and entertainment tagging.

Secondly, other researchers might extend the sample sizes to examine the process of tagging behavior through talk aloud experiments. The sample might focus on informants who know tagging actually rather than new taggers. Moreover, the informants should tags for their own languages rather than others since it would be natural tagging.

REFERENCES

Ames, M. & Naaman, M. (2007). Why we tag: Motivations for annotation in mobile and online media. In *Proceedings of the 2007 SIGCHI Conference on Human Factors in Computing Systems* (New York: ACM Press), 971-980.

Angus Beaton, Steven Nicholson, Neil Halliday & Keith Thomas (2010). Think-Aloud Protocols. Lectures number 5. Retrieved on 03.12.2010 from http://www.psy.gla.ac.uk/~steve/HCI/cscln/trail1/Lecture5.html

Bar-Ilan, J., Zhitomirsky-Ge_et, M., Miller, Y., and Shoham, S. (2010). The effects of background information and social interaction on image tagging. *Journal of the American Society for Information Science and Technology*, 61(5):940-951.

Bischoff, K.; Firan, C. S.; Nejdl, W. and Paiu, R. (2008). Can all tags be used for search? In *CIKM '08: Proceeding of the 17th ACM conference on Information and knowledge management*, 193–202.

Cattuto, C., Baldassarri, A., Servedio, V. D. P., and Loreto, V. (2007). Vocabulary growth in collaborative tagging systems. Retrieved on 29.11.2010 from http://arxiv.org/abs/0704.3316

Chi, E. H. and Mytkowicz, T. (2007). Understanding Navigability of Social Tagging Systems. *In Computer/Human Interaction*.

Dong, W., & Fu, W. (2010). Cultural difference in image tagging. In Proceedings of the ACM conference on computer-human interaction, 981-984. Atlanta: ACM

Golder, S. & Huberman, B. (2005). The structure of collaborative tagging systems. *Information Dynamics Labs*. Retrieve on 28th 10, 2010 from http://arxiv.org/ftp/cs/papers/0508/ 0508082.pdf.

Golder, S.A. and Huberman, B.A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science,* 32(2), 198-208.

Guy, M., Tonkin, E. (2006). Tidying up tags, *D-Lib Magazine*, 12(1).

Hammond, T., Hannay, T., Lund, B. and Scott, J. (2005). Social bookmarking tools: A general review. *D-Lib Magazine*, 11(4), 181-190.

ISO. (1985). *Documentation Methods for Examining Documents, Determining their Subjects and Selecting Indexing Terms*. International Organization for Standardization.

Kipp, Margaret E.I. (2007). @toread and Cool: Tagging for Time, Task and Emotion. In Proceedings of 8th Information Architecture Summit. Las Vegas, Nevada, USA. Retrieved September 30, 2007 from Web site: http://eprints.rclis.org/archive/00010445/

Korner, C. (2009). Understanding the motivation behind tagging. *ACM Student Research Competition-Hypertext.*

Kuhlthau, C. C. (2004). Seeking meaning: A process approach to library and information services. Norwood, NJ: Ablex.

Lee, C.S., Goh, D., Khasfariyati, R., & Chua, A. (2009). Tagging, Sharing and the Influence of Personal Experience, Journal of Digital Information, 10(1), 275.

Marvasti, A. F. (2008). Social structure in tagging practices: reality or myth? Kingston, Ontario, Canada: Queen's University.

Mathes, A. (2004). *Folksonomies - cooperative classification and communication through shared metadata*. Retrieved on November 3, 2007, from http://adammathes.com/academic/computer-mediated communication/folksonomies.html

Mitchell, Joan S., Julianne Beall, Winton E. Matthews, Jr., and Gregrory R. New, eds. (1996). *Dewey Decimal Classification and Relative Index. Devised by Melvil Dewey. Edition 21.* Albany: Forest Press.

Nielsen, J. (2000). Why you only need to test with five users, Jakob Nielsen's Alertbox, http://www.useit.com/alertbox/20000319.html

Nov, O., Naaman, M., and Ye, C. (2008). What drives content tagging: the case of photos on Flickr. *In CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, 1097-1100, New York, NY, USA, ACM.

Rashmisinha (2005). A cognitive analysis of tagging [Web log message]. Retrieved on 29.11.2010 from http://rashmisinha.com/2005/09/27/a-cognitive-analysis-of-tagging/.

Sen, S., Lam, S., Rashid, A.M., Cosley, D., Frankowski, D., Osterhouse, J., Harper, F.M., and Riedl, J. (2006).Tagging, communities, vocabulary, evolution. Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW 2006), 181-190.

Sinclair, J. & Cardew-Hall, M. (2008). The folksonomy tag cloud: when is it useful? Journal of Information Science, 34(1), 15-29.

Spitieri, L. (2007). Structure and form of folksonomy tags: The road to the public library catalogue. Information Technology and Libraries, 4(2), 459-467.

Sterken, V. (2008). User generated metadata in the LibraryThing folksonomy. *Vrije Universiteit Brussel*. Retrieve on 17th, November, 2010 from http://www.ethesis.net/folksonomy/folksonomy.pdf.

Stoyanovich, J., Yahia, S. A., Marlow, C., Yu, C., (2008). Leveraging Tagging to Model User Interests in Del.icio.us., In Proceeding of the AAAI Spring Symposium on Social Information Processing.

Strohmaier, M., Korner, *C. and Kern, R. (2010).* Why do Users Tag? Detecting Users' Motivation for Tagging in Social Tagging Systems*. In 4th International AAAI Conference on Weblogs and Social Media (ICWSM2010), Washington, DC, USA.*

Van Someren, M. W., Barnard, Y. F., and Sandberg, J. A. C. (1994). The Think Aloud Method: A Practical Guide to Modeling Cognitive Processes. Academic Press Limited, 218 p.

Vuorikari, R., Ochoa, X., and Duval, E., (2009). Analysis of User Behavior on Multilingual Tagging of Learning Resources. In *Proceedings of the 1st Workshop on Social Information Retrieval for Technology-Enhanced Learning & Exchange, (6-17).*

Wales, J. (2010). Cognitive psychology. Wikipedia. Retrieved from http://en.wikipedia.org/wiki/Cognitive_psychology

Wu, Chao and Zhou, Bo (2009). Analysis of Tag within Online Social Networks. *ACM digital library*, 21-30. Retrieved on 22.10.2010 from DOI: 978-1-60558-500-0/09/05.

Yin, D., Xue, Z., Hong, L., and Davison, B. D. (2010). A probabilistic model for personalized tag prediction. *In KDD '10: Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, 959-968. ACM: New York, USA.

APPENDIX 1: TAGGING TASKS DESIGN

1. *Task 1: Tagging in CiteULike*

From LibraryThing, choose and search *a topic you like* and then collect 1 possible document at the same with that topic. Then give tags for chosen documents, and while you do so, try to say everything that goes through your mind.

2. *Task 2: Full text tagging*

You are working in a Sex Safety Project which aims to help youths become aware of how to protect their sexual health. Your responsibility in that project is collecting related documents. There is an article "*Sex safety: the best protection against sexually transmissible infections, including HIV/AIDS*" in the laptop desktop. Please do anything you wish (you can use internet to check information if you want) to tag it in a paper and while you do so, try to say everything that goes through your mind.

3. *Task 3: Fictional tagging in LibraryThing*

Please choose one of the books following the table:

Book	Title 1	Title 2	Author
1	Harry Potter and the Philosopher's Stone	Harry Potter and the Sorcerer's Stone	J. K. Rowling
2	Where Rainbows End	Rosie Dunne	Cecelia Ahern
3	The Iron Man	The Iron Giant	Ted Hughes
4	Little Mexican	Young Archimedes	Aldous Huxley
5	Miss Smilla's Feeling for Snow	Smilla's Sense of Snow	Peter Høeg
6	Northern Lights	The Golden Compass	Philip Pullman
7	The War of Dreams	The Infernal Desire Machines of Doctor Hoffman	Angela Carter

Then, search for both titles of that book and give tags for them. When doing the task, please try to say anything that goes through your mind.

4. ***Task 4: Video tagging in YouTube:*** Pick any video on YouTube of your choice and tag it in paper, while doing so; try to say anything that goes through your mind.
5. ***Task 5: Image tagging in Flickr:*** Pick any picture on Flickr of your choice and tag it in paper, while doing so; try to say anything that goes through your mind.

APPENDIX 2: GUIDELINE OF TALK ALOUD EXPERIMENT

Dear friends,

First of all, I would like to say many thanks to you.

I am happy as you are here for the talk aloud experiment which aim to know tagging process from diverse people; to understand behavior in each tagging process and hopefully I may figure out tagging troubles to be supported from systems.

To do the talk aloud experiment, I am going to give some guidelines as follows:

1. Talk aloud experiment:

- The test will give you 5 tasks to do.
- During do the task; please try to say anything goes through your mind. Then keep talking to describe what you are thinking and feeling.
- For example, "*I am feel uncomfortable do tag this object as I don't know much about topics and …*"

2. LibraryThing, YouTube and Flickr are experimental sites:

LibraryThing uses to tag for books, articles tasks whereas YouTube and Flickr tag for video and image tasks.

You will have 10 minutes for each site to get acquainted with:

- The general structure of websites.
- Understand how to search, navigate and do tag in each site.

If there are any questions about this, please let me know to help you.

3. Structure of the experiment:

- Time to take test: approximately 1 and a half hours
- Process of test:
  - Pre-test : 5 survey questions

    - During-experiment : 5 tasks (talk aloud to do them)
    - Post-experiment : 5-8 interview questions.
- Method to do experiment:
    - Follow and do tasks in convenience as you do it in nature.
    - There is no time limitation, since this work needs time for both think and talk.
    - Try to say anything goes through your mind when doing each task.
- Questions:
    - Participants are free to ask moderator (me) for any questions to be clear about tasks, method to do experiment or technical problems arise when tagging.
    - To be positive results, some questions about how to search or tag for objects are out of responsibilities of moderator.
    - The questions and answers should be given after finishing the talks of participants, as they could disturb participants' though.
    - As talk aloud experiment, in case of participants give up talk louder while tagging, moderator can ask to encourage them keep on think aloud.

4. Languages to tags:

Entire objects and tasks are written in English, moreover, the sites also English cites so that the tags are expected in English.

APPENDIX 3: PREPARTIONS FOR THE EXPERIMENT

1. Prepare a topic of your own choice to use for tagging a book in LibraryThing. The topic can be, e.g., academic topics, fiction, or something else within your interest sphere. We will use it in the experiment.

2. Please prepare to choose one of the following books for tagging in LibraryThing. Each of the books has two different titles but they are still at the same book:

Book	**Title 1**	**Title 2**	**Author**
1.	Harry Potter and the Philosopher's Stone	Harry Potter and the Sorcerer's Stone	J. K. Rowling
2.	Where Rainbows End	Rosie Dunne	Cecelia Ahern
3.	The Iron Man	The Iron Giant	Ted Hughes
4.	Little Mexican	Young Archimedes	Aldous Huxley
5.	Miss Smilla's Feeling for Snow	Smilla's Sense of Snow	Peter Høeg
6.	Northern Lights	The Golden Compass	Philip Pullman
7.	The War of Dreams	The Infernal Desire Machines of Doctor Hoffman	Angela Carter
8.			

3. Please prepare to tag a picture of your own choice in Flickr.

4. Please prepare to tag a video of your own choice in Flickr/YouTube.

APPENDIX 4: QUESTIONNAIRES FOR THE PRE-EXPERIMENT

In order to have clear information about your experience on tagging, please fill up and answer short questions as follows:

1. Your level of tagging experience?

Excellent	Good	Fair	Non-skill

2. Have you ever done tagging in LibraryThing?

Yes	No

3. Have you ever done tagging in YouTube?

Yes	No

4. Have you ever done tagging in Flickr?

Yes	No

5. Do you have much knowledge on each topic of tasks:

Topics	Excellent	Good	Fair	Non-knowledge
Task 1				
Task 2: Sex safety				
Task 3				
Video task				
Image task				

6. Please give some information about yourself:

- Name:
- Sex:
- Job before come with DILL:
- Education:

It's time to do talk louder experiment!

APPENDIX 5: INTERVIEW QUESTIONS

1. What are important things for looking before tag?

2. Before type out the tags, what you're thinking and feeling at that time?

3. Do you want to give many tags in each document?

4. Questions for each task:
   - How do you feel to tag in task 1, 2, 3, 4 and 5?
   - Which task you meet difficulties? Why?
   - Which task you feel comfortable? Why?
   - Explanations about tagging results:
     - Task 1:
     - Task 2:
     - Task 3:
     - Task 4:

5. Evaluation about tagging systems:
   - Ease to tag?
   - Offer functions to support

6. Some suggestions for each tagging systems

7. There are some additional questions from the observation:

APPENDIX 6: TAG RESULTS

Article in CiteULike:		
Types of tag	**Participants**	**Tags result**
Topics	#R	Otaku, videogames, playercreated content, neofiction, user-behaviuor
	#A	Brasov history
	#L	topic maps, information organization, CEDECA project, cultural heritage, Aquabrowser, (libraries, archives, museums) LAM
	#Jen	Kingdoms, monarchies, trilogies, Fantasy
	#E	Moses and the ten commandments, The birth of Jesus, The story of Joseph in Egyp, Noah ark
	#J	cyberpunk, utopia, dystopia
	#K	political communication, habermas, deliberation, legitimation, media societies
	#C	cosmetics, skin, consumer, protection, products, ingredients
	#W	Hamlet, adaption, drama
Author	#J	William Gibson
Type of document	#J	science fiction
	#W	theme, theory

Sex safety booklet:		
Types of tag	**Participants**	**Tags result**
Topics	#R	HIV, AIDS, STI, Sexually Transmissible Infections, Condoms, Herpes, Genital wants, Chlamydia, Gonorrhoea, Trichomonas vaginalis, Variatious
	#A	Youth sex education
	#L	Safe sex, Sexual transmissible infection (STIs), HIV/AIDS, Sexual health, Sex protection
	#Jen	STI, HIV, AIDS, Contraceptives, Condoms.
	#E	Safety sex, condoms, sex, Protected sex
	#J	Sex, education, HIV, AIDS
	#K	Safe sex, transmissable diseases, HIV, clinics, protection, STI clinics, symptoms, treatments, risks, information strans dis
	#C	Sex, protection, safe sex, desease, , condoms, oral sex, , STI, HIV, youth, help, guide, help,
	#W	safe sex, STI, condoms, STI&HIV
Task organization	#C	Definition, Victoria Government Dpt, Information,

Fictions /participants	**Types of tag**	**Tags result**
Miss Smilla's Felling for Snow/#R	Topics	crime fiction, snow, Iniut, Miss Smilla's Felling for Snow
	Locations	Greenland, Denmark
	Author	Peter Hoeg
	Types of document	children fiction
	Personal idea	Suspense, Sad
#W	Topics	snow, children
	Types of document	detective story
	Time	post-colonization
Northern Light/ #A	Topics	Adventure
	Task organization	his dark materials
Harry potter and the philosopher's stone/ #L	Topics	wizards, magic, Witchcraft
	Types of document	made into movie
	Personal opinion	best-selling fiction
#Jen	Topics	school, wizards, witches, magic, children, fantasy
	Types of document	Fiction
#E	Topics	Professor Dumble Doe, Hogwarts School, Harry Porter, Lord Voldemot, Ron Weasley, Wizards
	Author	J.K.Rowlling
	Types of document	British fiction
#J	Topics	young adults, magic fantasy
	Types of document	Fiction
#K	Topics	harry potter, Hogwarts, dumbledore, voldemort, philosopher stone, ron weasley, hermioni granger, children, adults, the deadly hollows
	Types of document	the movie, fantasy, magic
	Author	j.k.rowling
	Task organization	Volume
#C	Topics	Harry Potter, wizard, owl, magic, kids, children, Sorcers's stone, Philosopher's stone

Video:	
Types of tag	**Tags result**
Topics	Murderer, Prison-Rehabilitation, traels(Annoying), Tourist, Festivals, Naïve advertisement, selling Colombia, birds, ducks, ducklings, wind, mom, When I look at you, The Last song, Liam Harmsworth, Miley Cyrus 2010 songs, Santa, Esmeralda, disco, Flamenco, Salsa, exotic, dancing, steps, history, basic steps, Lush, crush, bobbles, bath, slice, soap, bobble bath, pink star, battle baub, guitar, love story, theatre, Romeo and Juliet, Hannah Montana
Locations	Colombia, Latin American, amsterdam
People	Miley Cyrus
Title	don't let me be miss-understood,
Time	Miley Cyrus 2010 songs, 1977
Types of documents	Interview, music video, music
Personal idea	Idiot, Optimistic, Colombia is passion, funny, Love/inspirational song, Uplifting, Troubled teenager, Young love, Childhood innocence, famous music, popularity, famous dancer
Task organization	Kaffe med Kunrt (Coffee with Kurt), Forvalsadvokaterne.com, , salsa for advanced dancers

Image:	
Types of tag	**Tags result**
Topics	The Globe, The tourist, Midnight Sun, London eye by night, library park, Medellin, Biblioteca Espana, Social project, community development, Libraries, Churches, Sunny, Baltics, architecture, cathedral, Churches, Toompea, Winter season, Christmas season, DILL 3, rockets, sear waves, sun, tunnel, sur, canal, boat, tour, city, spring, summer, central station, waterhouse, bus, bines, craine, pier, Geisha, laugh, , Kimono, make up, Chinese lantern, ; sea, green plants, houses
Locations	Nordkap (North Cape), North Cape, "Northern-Most-Part-Of-Europe", Colombia, Sergio Fajardo, Giancarh Mazzanti (architecture), Tallinn, Old Town, Estonia, Conad supermarket, California, Pfeiffer Beach, Asia, Japan, Ireland, countryside
People	Ezerea, Andrew, Yalda
Personal idea	Cold, Scenic, big, fun
Task organization	To-do-list

APPENDIX 7: TOPIC EXPERIENCE

FACTUAL TOPICS			
Participant	**Chosen topics/documents**	**Experience**	**Explanation**
#R	Playing with Videogames	-Excellent -Non-reading	-Interesting topic
	Sex Safety: the best protection against…	-Excellent -Non-reading	-Educated topics at schools
#J	Cyberpunk: science fiction	-Excellent -Non-reading	-Interesting topic
	Sex Safety: the best protection against…	-Excellent -Non-reading	-Educated topics at schools
#A	Thermal analysis of Romanian ancient ceramics	-Good -Non-reading	-Her home country
	Sex Safety: the best protection against…	-Good -Non-reading	Educated topic but she didn't like it too much.
#L	Navigating Through archives…	-Excellent -Non-reading	-She was researcher in this. -She hasn't read the article but she knew topic clearly.
	Sex Safety: the best protection against…	-Good -Non-reading	-Educated topic in school -Interesting topic
#Jen	What the batman think about SpongeBob…	-Fair -Non-reading	-Interesting topiC
	Sex Safety: the best protection against…	-Good -Non-reading	-Educated topic in school -Interesting topic
#E	Bible stories for children	-Excellent -Read book	-Interesting topic -She read it long time ago
	Sex Safety: the best protection against…	-Good -Non-reading	-Educated topic in school -Interesting topic
#K	"Political communication and the Epistemic…"	-Good -Non-reading	- Her interesting research
	Sex Safety: the best protection against…	-Good -Non-reading	-Educated topic in school -Interesting topic
#C	Don't go to the cosmetics counter without me…	-Good -Non-reading	-Researching in this field -She never read it before
	Sex Safety: the best protection against…	-Good -Non-reading	-Educated topic at school
#W	Revenge, Honor, and Conscience in "Hamlet	-Fair -Non-reading	-Her interesting research topic
	Sex Safety: the best protection against…	-Good -Non-reading	-Educated topic at school

FICTIONAL TOPICS			
Participants	**Chosen topic and document**	**Experiences**	**Explanations**
#R	Smilla's Sense of Snow	Good Read book	-Danish fiction -He read it for long time so that his memory of the details wasn't clear.
#W	Smilla's Sense of Snow	Fair Non-reading	-She just knew in general
#A	The Golden Compass by Philip Pullman	Good Read book	-She didn't know The Golden Compass and "Northern Light" is the same book or not.
#J	Harry Potter and the philosopher's stone	Fair Non-reading	-He knew it in general
#L	Harry potter and the philosopher's stone	Fair Non-reading	-She knew it in general
#E	Harry Potter and Philosopher's stone	Fair Non-reading	-She knew through media -She has general understand about topic such as names of main players, author as well.
#Jen	Harry potter and the philosopher's stone	Good Read book	-She knew both titles are the same book.
#K	Harry Potter and Philosopher's stone	Good Read book	-She read it before
#C	Harry Potter and the philosopher's stone	Good Non-reading	-She knew through media and friend -She has general understand about topic such as names of main players, author as well.

ENTERTAINMENT TOPICS			
Participants	**Chosen documents**	**Experiences**	**Explanations**
#R	Video: "han havde været træls"	Excellent Watched	-He watched it before
	Image: "Nordkap"	Good Unwatched	-He has known -He has never seen
#J	Song: Santa Esmeralda	Excellent Watched	-He knew it before
	Image: Big Sur	Good Unwatched	-But he knew the place well
#A	Video: Trailer film	Excellent Watched	-Interesting topic
	Image: London	Excellent Watched	-She took that picture
#L	Video: Columbia	Excellent Watched	-Her home country
	Image: Medellin Library	Good Unwatched	-Her working place
#Jen	Video: Duck blown…	Excellent	-Interesting video

		Watched	
	Image: Old town in Tallin	Excellent Unwatched	-Interesting image -She knew the place well
#E	Video: song of Miley	Excellent Watched	-Interesting topic
	Image: her picture	Excellent	-Interesting topic
#K	Video: Salsa music	Fair Unwatched	-Interesting topic -She knew everything in the video
	Image: Amsterdam	Excellent	-Her home
#C	Video: "LUSH product .."	Good Unwatched	-Her researching field
	Image: 3 Japanese girls	Good Unwatched	-Good knowledge
#W	Video: Romeo and Juliet	Good Unwatched	-Interesting topic
	Image: Howth and Ireland's Eye	Fair Unwatched	-Interesting topic -Never seen before