Tags and Information Recollection

Max Chevalier
Institut de Recherche en Informatique de Toulouse (IRIT),
UMR5505 CNRS,
IUT Paul Sabatier,
Université de Toulouse,
France
max.chevalier@irit.fr

Josiane Mothe
Institut de Recherche en Informatique de Toulouse (IRIT),
UMR5505 CNRS,
Ecole supérieure du professorat et de l'éducation (ESPE),
Université de Toulouse,
France
josiane.mothe@irit.fr

Patrice Terrier
Laboratoire Cognition, Langues, Langage, Ergonomie (CLLE),
UMR5263,
Université de Toulouse,
CNRS, UT2J,
France
terrier@univ-tlse2.fr

***Abstract*—Folksonomy, tag, bookmark and social bookmarking are all concepts that refer to recollection in information retrieval (IR). Recollection can be defined as the effort needed to recall an even, an idea or a resource. In IR, when locating an interesting resource, a user can tag it in order to "memorize" it and retrieve it later. The user chooses the tags according to his/her needs or centers of interest. While previous work has addressed the problem of tag exploitation, to our knowledge, an important dimension is missing. None has focused on the tagging activity itself when considering the users point of view. An important open question is the following: will a user be able to remember tags he/she associated with a document in the past in order to retrieve a document again? Our paper tackles this problem. Considering a user study that implied 32 participants, we evaluate some elements (age, time...) that impacts the recognition and the recall of tags used in the past. This result is important specifically as it can be used in order to optimize personalized systems (e.g. user profiling, tag exploitation...).**

I. Introduction

This paper addresses the issue of recollection in information retrieval. We focus on tagging, which is a major mean for recollection for electronic data, specifically on the web. Tagging is popular on the web because it allows a user to share and retrieve web resources. On an online social bookmarking platform [1], tags are organized into a folksonomy. In this paper we provide details of an experiment we designed to determine particularly the impact of human characteristics on tagging and the resulting recollection. The user study implied 32 participants. Three main dimensions have been identified and evaluated (effects of the studied characteristics on the selection of terms/tags, on the recall of tags and on the recognition of tags). The results of the study are examined in detail. This study can be useful for instance when designing online systems e.g. collective, adaptive or personalized systems.

The paper is set out as follows: section II discusses related work and defines the concepts of tagging and recollection. It also highlights the relationships between recollection and tagging as discovered in previous work. Section III introduces the study we designed to determine the effects of human characteristics (age, knowledge/comprehension, memory/time, intent) on tagging. Section IV contains the main results. The last section consists of the conclusion and recommendations for future work.

II. Tags for recollection: definitions and related work

A. Definition: tag

In this paper a tag is defined as a "label". It contains either a single word or a bag of words. Thus, a "tag" does not refer to an HTML tag, e.g. <p>, nor to an HTML metadata element (used in the HTML head section). One important dimension of tags is that they are generally defined by end-users without reference to a particular dictionary or ontology.

To tag a resource the user has to associate one or more tags i.e. word(s) with it. It is important to note that each user can associate different tags with the same resource, as tags interpret the users view of this resource. So for a user, a tag can correspond to a category, a project or a center of interest that he/she considers as being connected to the resource.

B. Definition: recollection

Recollection is an important dimension in todays online systems or platforms: the web abounds with resources (texts, images, videos, etc.). For instance, in May 2016, the size of the indexed World Wide Web was at least 4.58 billion pages[1]. The indexed web corresponds to the content that can be indexed by a search engine like Google. As a consequence, a user is submerged in information he/she cannot reasonably remember. At best he/she may have a vague memory of resources that have been encountered.

Recollection corresponds to the way a user remembers a resource (its content, the reason why he/she encountered this resource earlier or the reason why he/she used this resource, the website on which the resource was available, the context in which he/she found this resource, etc.).

In this paper, we only consider the online information seeking dimension of the recollection, even if recollection can also be applied to offline systems (e.g. in common word

[1]http://www.worldwidewebsize.com/ (visited may 11th 2016)

processors software). We examine how a user is able to retrieve/remember an online resource that he/she has already found [2]. Information seeking consists of looking for a resource or information using any available means (reading a book, asking a friend, choosing an online search engine, etc.).

Before 2000, returning to a previously located web resource was facilitated by what were called ”bookmarks” [3]. Bookmarking is a function offered by any web browser. The limitation of bookmarks is that they must be structured as a set of directories. Such a structure (corresponding to a tree) can only store a small number of bookmarks because it becomes too difficult to manually manage and maintain a high number of hierarchically organized bookmarks. Indeed, a directory containing a lot of bookmarks should be split into sub-directories to keep the tree coherent. Bookmark management was considered by users as boring and a very time consuming task.

Using bookmarks is however an interesting way to help a user retrieving resources, since they are organized according to the users point of view. Indeed, a bookmark directory tree corresponds to the users vision of all the resources. Another limitation is the difficulty in manually adding a resource to several directories. Lastly, bookmarking, having only been developed within web browsers, has met with scant success. Indeed, a bookmark only contains the URL of the resource and clicking on a bookmark opens the resource in the browser. Even if various systems like [4] have tried to help users manage and exploit their bookmarks to improve the search experience, bookmarks have been increasingly ignored by users. We have to point out that bookmarking is still available in todays web browser, but with the same limitations.

In early 2000, tagging appeared and was an attempt to solve some of the limitations associated with using bookmarks. The outcome of tagging is a flat organization of words that is to say there is no hierarchical organization. Indeed, after a user defines several tags to describe a resource, he/she may, within this list of tags, interpret the various views he/she holds on this resource. For instance, when tagging the resource https://en.wikipedia.org/wiki/Java, one may associate tags like ”geography”, ”island”, ”tsunami” or ”holidays” that cover different relations between the user and the tagged resource. Bookmarking the same resource would imply adding this resource (its URL) to several directories in the bookmark tree, each one corresponding to a specific point of view.

Here we should note that tagging is available in many online and offline systems, such as word processors, operating systems, etc. For the purposes of recollection, bookmarking and tagging facilitate the retrieval of a previously located resource. Due to the limitations of bookmarks, in this paper we only focus on tagging as a tool to improve the recollection of online resources.

C. Tags for Recollection

In recent years, tagging has become a standard function available on many online systems. Thus, we can tag almost any resource (regardless of type and format) via a number of online systems and platforms, such as *Instagram* and *Delicious*. The reasons for the spread of tagging are essentially as follows:

- Tagging appears for the users an easy activity (if we do not consider the relevance of the chosen tags) [5] and does not require any additional knowledge. The activity is suitable for almost any user;
- Clicking on a tag allows a user accessing to every resource which has been tagged with a particular tag, so relationships between various resources can be established/discovered via the tags that have been shared between these resources i.e. the folksonomy;
- The growth of collective or collaborative activities. Thus, with collaborative platforms like *Delicious* [2] and *BibSonomy*[3] (set up for scientific publications and resources), users can share their tags and the associated resources.

Fig. 1. A tag cloud

The last two reasons have paved the way for users to gain access (through their own tags) to additional resources that have been retrieved by other users, hence tagging combines social and individual dimensions [6]. Note that this can only be of interest if users have the same vocabulary. As a solution, users can search for resources through querying tags for such platforms. Another possibility is to display a set of available tags in a tag cloud (Figure 1). The user could simply click on a displayed tag to see related resources. This kind of visualization facilitates information retrieval and recollection via the available tags [7], [8]. Moreover, tags have been intensively studied for their application to various domains: annotation systems [9], information retrieval [10] (for music [11], video [12], etc.), recommender systems [13].

Tagging has also been studied for tagging patterns [14], user profile generation [15], [16] and optimum tag selection (particularly with regard to folksonomies [17]).

As we can see, tags are important clues for recollection, but to our knowledge in previous studies the link between recollection and tags has essentially been examined with regard to how systems exploit tags. In this paper we study recollection in terms of the user dimension in tagging. We designed an experiment that attempts to analyze the effects of human characteristics on tagging.

[2]http://del.icio.us/

[3]http://www.bibsonomy.org/

III. Studying the Impact of Human Characteristics on Recollection

A. Objectives

The objectives of this study were to analyze the effects of human characteristics on recollection. To remain as close as possible to a real case, participants were asked to tag a web resource. Before giving details of the various hypotheses and the participants, we list the characteristics we considered for this study:

- age: how old is the user?
- knowledge and comprehension levels: does the participant know and understand the proposed documents?
- memory/time: does the participant forget the tags he/she associated with a resource?
- intent: does the user want to share his/her tags [public tags] or not [private tags]?

The different hypotheses account for the impact of these characteristics on the number of tags chosen and the way they are chosen (inside or outside the document), the recall and the recognition values (these values are defined in the next section).

After examining these effects, some preliminary recommendations about tag usage for the recollection of web information can be done. For instance, when profiling users for information retrieval, system mainly exploit terms that are present in tagged documents. Is it is enough or should we consider some external terms in a more important way according to the user characteristics in the profile?

The following section discusses the hypotheses and the corresponding analysis protocol we designed.

Please note that this study only considers textual documents, such as web pages.

B. Experimental Protocol

We first defined two different categories of users according to their age: U1 (18-35 years old) and U2 (58+ years old) in order to find out if age has an effect on recollection (in order to have a difference high enough between these two categories). In order to identify the impact of knowledge on recollection and to ensure that we had a wide range of knowledge/comprehension levels, we decided to use:

- two "*unkown*" documents (low knowledge level). We chose two topics for which most people are expected to have low knowledge and comprehension levels: one document related to nanotechnologies and the other to astrophysics;
- two "*known*" documents (high knowledge level). To be sure that each participant really knew and understood at least 2 documents, they provided their own documents. Therefore, the sets of documents were likely to be different from one participant to the others.

Moreover, our protocol was based on three time phases: T0, T1 (T0+3 days) and T2 (T0+7 days), so that we could study the effect of time on human (long-term) memory.

Lastly, to be sure that the analysis covered all cases, and to enable the examination of the obtained results, a specific distribution of tasks for each participant was predefined.

Figure 2 shows the distribution of participants (subjects) and the corresponding tasks. In this figure, Text #1 and Text #2 correspond to "unknown" documents while Text #3 and Text #4 correspond to "known" documents.

Details of the time phases:

- T0: "*tagging*": At T0, a participant had to tag the two unknown documents according to the given intent (visibility: public/private). After, he/she gave the two URLs corresponding to documents that he/she really knew and understood. He/she also had to define the tags he/she wanted to associate with these two known documents, and maintain the same visibility for the unknown documents.
- T1, T2: "*remembering*": At T1 and T2, each participant had to remember the tags he/she associated at T0 with two specific documents (one is known; one is unknown - see 3.2). This is the core of our study because it allows us to see the impact of time on human memory and recollection. To measure this, we used common memory tests from psychology: Recall and Recognition test [18], [19].

  To evaluate "Recall", we asked each participant to write all the tags he/she could remember for a given document. The evaluation of the recall was carried out as follows: we assumed that a participant associates n different tags with a document at T0. We then measured the number of correct tags (n^{Correct}) at T1 and T2 which corresponded to the tags each participant actually used at T0. In the same way, we measured the number of incorrect tags ($n^{\text{Incorrect}}$) which did not correspond to the tags each participant used at T0. From this we identified three possibilities for measuring the recall value: common (Equation 1), global (Equation 2) and complete recall (Equation 3).

$$commonRecall = \frac{n^{Correct}}{n^{Correct} + n^{Incorrect}} \quad (1)$$

$$GlobalRecall = \frac{n^{Correct}}{n} \quad (2)$$

$$completeRecall = \frac{n^{Correct} - n^{Incorrect}}{n} \quad (3)$$

  In order to precisely measure recall we used in this paper the Complete Recall formula (Equation 3). It computes the strength of recall by moderating the result according to the number of incorrect tags. One may consider that it should reflect the worst case.

  To evaluate "Recognition", we asked each participant to identify the words that he/she really used at T0 from a list of words (randomly ordered words that the participant used as a tag for the given document were combined with "distractor words" displayed in a single column. Half of the distractor words have been chosen inside the document content, half have been chosen outside

Visibility	Age	Subject	Phase #1				Phase #2			
			Text #1		Texte #2		Texte #1		Text #2	
			Text #3	Text #4	Text #3	Text #4	Text #3	Text #4	Text #3	Text #4
Public	[18-35]	S1-S2	x							x
		S3-S4		x					x	
		S5-S6			x			x		
		S7-S8				x	x			
	[> 58]	S9-S10	x							x
		S11-S12		x					x	
		S13-S14			x			x		
		S15-S16				x	x			
Private	[18-35]	S17-S18	x							x
		S19-S20		x					x	
		S21-S22			x			x		
		S23-S24				x	x			
	[> 58]	S25-S25	x							x
		S27-S28		x					x	
		S29-S30			x			x		
		S31-S32				x	x			

Fig. 2. Task/users distribution.Text#1 and Text#2 are "unknown" whereas Text#3 and Text#4 are "known". The visibility corresponds to public/private tags.

the document content. Similarly to recall, the recognition value can be computed in 3 ways: common, global and complete recognition. Both the recognition formula and the recall formula are based on the same procedure. As with recall, we analyzed the complete recognition (Equation 4).

$$completeRecognition = \frac{n^{Correct} - n^{Incorrect}}{n} \quad (4)$$

C. Analysis methods

We used statistical methods on the results of the experiment to verify the correlation, the significance of the difference between the results and so on. These statistics were computed with IBMs SPSS software. To avoid being too verbose, we have cited references only at the first instance of a test. Without additional precision, and adhering to the assumption of normality (not provided for each assumption), the adapted test was selected. Moreover, any given significance value is exact (not asymptotic) and is based on a 2-tailed significance test.

Note that due to the page limit we report the most relevant results only.

TABLE I
PARTICIPANT AGE

Age Category	Nb Part.	Mean Age	Std Dev.
18-35	16	26.2	5.05
58+	16	63.8	4.04
All	*32*	*45*	*19.6*

TABLE II
PARTICIPANT EDUCATION LEVEL (20 = PHD)

Age Category	Mean	Std Dev.
18-35	17.25	1.54
58+	16.29	3.61

TABLE III
PARTICIPANT INFORMATION RETRIEVAL SKILL LEVEL (0-WORST ; 10-BEST)

Age Category	Mean	Std Dev.
18-35	7.125	1.54
58+	7.31	1.74

Value/Scale	Education Level
1	None
9	8th grade (4ème)
13	12th grade (BAC)
16	Bachelor (Licence)
20	PhD

Fig. 3. Education Level Scale

D. Participants and Document Collection

We asked 32 people to participate in the experiment. To recruit them a call for participation was spread via mailing-lists. All participants achieved their task (since they received a gift coupon when finishing the experiment). Table I shows the mean age for the 16 users in the [18-35] category which is 26.2 years old (SD[4]=5.05), while the mean age for the 16 users in the [58+] category is 63.8 years old (SD = 4.04). Table II completes the characteristics with participants' education level (value 20 corresponds to a phD level) see Figure 3.

Each participant, after completing the experiment (around 25 minutes for the three phases: T0, T1 and T2) received a gift coupon. To complete the experiment, a tool (based on a client-server architecture) was implemented. This tool enabled the user to achieve his/her task and to submit his/her two known documents at T0. In an effort to achieve a more thorough analysis and verification, this tool stores the content of each document given by a participant i.e. the known document in HTML and PDF format. Moreover, it gives additional variables like the duration of each phase for each participant. To carry out the best analysis as possible, we collected and compared the IR skill level for participants (see table III). This value (that corresponds to their familiarity with internet) was evaluated by means of a questionnaire. All the participants had been using internet for at least 4 years (frequency of use: at least 2 hours per week). This IR skill level takes account of the frequency and diversity of internet services used [0 (low); 10 (high)]. It is interesting to note that the two age categories have a similar mean and standard deviation for the IR skill level ([18-35]: avg[5] = 7.125, SD = 1.54 - [58+]: avg = 7.31, SD = 1.74). This is perhaps surprising as we could have hypothesized that participants belonging to the category [18-35] have a better IR skill level than those belonging to the category [58+]. This results from the recruitment for the experiment since participants have a good computing skill level. The topic of the study may have impacted the recruitment (since they may know what a tagging activity is). After a more thorough analysis of the data, a Mann-Whitney test indicated that the IR skill Level for [58+], IRLevel, (Mdn[6] = 8.0) was on the verge of being significantly different to the IR skill level for [18-35] (Mdn = 7.5), U = 78.5, p[7] = .05, r[8] = .35. Note: a value of around 0.3 for r (effect size) is considered as a medium effect size.

With regards to the knowledge and the comprehension of the documents, figure 4 shows the main statistics.

The knowledge and comprehension levels are given by each user for each document according to a "Likert scale" [0 (low); 10 (high)]. The last column shows the average time required by participants to process the two known documents and the two unknown documents at T0.

The knowledge and comprehension for known documents is significantly higher than for unknown documents ($p < 0.001$). Furthermore, there is no significant difference between the two age category values for knowledge and comprehension, nor a significant difference between the two visibility categories (private/public) for knowledge and comprehension. As we can see in Figure 4, the average experiment time for participants to process known documents (Mdn = 128,419.25) was significantly lower (more than 50%) than the average time required to process unknown documents (Mdn = 309,701.50) with Z = -4.6, $p < 0.001$. This result tends to suggest that users spent more time studying documents they knew little about (knowledge level).

IV. Main Results

This section presents the different hypotheses we verified and the results we obtained.

A. H1 - Selection of terms/tags

The first hypothesis we verified concerns the number of tags and the way they are chosen. Indeed, we study if words exist in the document content or not, that is to say if users choose words that are inside or outside the document they are tagging.

- H1.1. Effect of knowledge and comprehension levels - **partially validated**.
  There is a weak Kendalls tau correlation between the knowledge level and the proportion of tags chosen outside the documents: r=.125 with p=.024. At the same time, a weak negative correlation exists between the number of tags chosen at T0 and the proportion of tags chosen outside the documents r=-.158 with p=.033 i.e. the more the number of tags is high, the more the proportion of tags chosen inside the document is high.
- H1.2 Effect of intent i.e. the visibility of tags: public/private - **partially validated**.
  To verify this hypothesis, we used a Mann-Whitney test (2 independent samples) and compare the number of tags defined at T0 for private and public visibility. No significant difference was observed.
  We refined these results by analyzing the complementary effect of the knowledge level. We observed that for unknown documents, the proportion of tags chosen outside the documents after sharing tags (public) is significantly

[4]Standard Deviation
[5]Mean value
[6]Median
[7]p-value. A p-value lower or equals to 0.05 corresponds to strong presumption against null hypothesis.
[8]Correlation Coefficient

	Knowledge [0 ; 10]		Comprehension [0 ; 10]		T0 experiment time
	Mean	Std. Dev.	Mean	Std. Dev.	(1/1000 sec)
Known doc.	8.5	0.87	8.95	0.81	157371.78
Unknown doc.	2.516	1.91	7.07	1.15	353770.86

Fig. 4. Document related characteristics

lower than the proportion of tags chosen outside the documents when the tags are private (U=377, p=.022, r=.29).

Lastly, we also found that there exists an interesting correlation, which is significant (but weak) for the public visibility between the knowledge level and the proportion of tags chosen outside the text (r=0.227 with p=.024) i.e. after sharing tags (intent: public), the more the knowledge level is high, the more the proportion of words chosen outside the document is high (moderately).

- H1.3. Effect of age **validated**.
  With regards to the number of tags chosen at T0, we observed that the difference between the two age categories tends to be significant (p=.051). Indeed, we found that the number of tags defined at T0 for the [18-35] users (Mdn=4) is higher than the number of tags for 58+ users (Mdn=3).
  A Mann-Whitney test underlines the fact that there is a significant difference between the proportion of tags chosen outside the documents for the [18-35] category (Mdn=0) and the proportion of tags chosen outside the documents for the [58+] category (Mdn=.15) with p=.029.
  The proportion of tags chosen outside the documents is higher for users from the [58+] category than the proportion of tags chosen outside for users from the [18-35] category.
  When considering different combinations of visibility and age values for known and unknown documents, we found an average significant correlation (r=.572 with p=.009) between the number of tags used at T0 and the knowledge level for (58+/private/unknown documents).
  At the same time, for the same combination there is an average negative correlation between the knowledge level and the proportion of tags chosen outside the document, implying again that additional tags are essentially chosen within the document.

B. H2 - Recall of tags

Considering their definitions (cf. section 1), recall and recognition values have the following range: [-1; 1]. 1 indicates that a user recalls (recognizes) every tag he/she uses at T0. -1 indicates that a user recalls (recognizes) none of the tags he/she uses at T0 and supplies only incorrect tags.

The average recall for all participants (average value at T1 and T2) is equal to .297 (SD = .32) with a standard error equal to .056. These values indicate that the recall is relatively different between users. We then studied the effect of the human characteristics on the recall of tags at T1/T2.

- H2.1. Effect of knowledge and comprehension levels **validated**.
  Using the Wilcoxon Signed Ranks Test, we observed that:
  - On average (combining results for T1 and T2), the average recall for unknown documents (Mdn=.125) is lower than the average recall for known documents (Mdn=.51), Z=-2.451, p=.013;
  - At T1, the recall for unknown documents (Mdn=.24) is lower than the recall for known documents (Mdn=.48), Z=-1.949, p=.051 (tends to be significant);
  - At T2, there is no significant difference between the recall for known documents and the recall for unknown documents (p=.07).

  The recall value appears to be related to how much knowledge the participant has on the document content (even if participants use double the time to read unknown documents, which implies a longer encoding phase). This is confirmed by a (weak) correlation between the average recall and the knowledge level (r=.392 with p=.026).
- H2.2. Effect of intent (public/private) - **not validated**.
  Using a Mann-Whitney test we observed that there is no direct impact of visibility (public/private) on recall.
- H2.3. Effect of age - **partially validated**.
  We observed a significant difference (T-Test) (t^9=2.146, df^{10}=30, p=.04) between the average recall for users from the [18-35] category (Mdn=.47, Mean=.41, SD=.27) and the average recall for users from the [58+] category (Mdn=.19, Mean=.18, SD=.32).
  The average recall for younger participants is higher than the average recall for older participants.
- H2.4. Effect of Time - **not validated**.
  At first glance, time (T2 vs T1) does not have an impact on the recall value. Note that the delay between T1 and T2 is only 4 days. Maybe a longer period would lead to a more important effect. Notwithstanding, to avoid any bias related to the learning of the task we did not add a third phase (T3) for the same documents. An alternative to add a third phase T3 would be to renew the experiment with new documents. We carried out further analysis to see

[9]T-test value
[10]Degree of Freedom

if any other characteristics could moderate the result. No additional result was identified in this additional analysis.

C. H3- Recognition of tags

The average recognition for all participants (average value combining T1 and T2) is equal to .67 (SD = .24) with a standard error equal to .042.
The average recognition (Mdn=.72) is significantly higher than the average recall (Mdn=.31) for all participants Z=-4.843, $p < 0.001$. We then studied the effect of the human characteristics on the recognition of tags at T1/T2.

- H3.1. Effect of the knowledge and comprehension levels - **not validated**.
  Knowing or not knowing a document seems to have no impact on recognition.
- H3.2. Effect of the intent (public/private) - **not validated**.
  We did not find any effect for visibility on the average recognition value.
- H3.3. Effect of age - **not validated**.
  We did not find an effect for age on the average recognition value.
- H3.4. Effect of time - **partially validated**.
  An impact was found for the recognition value in relation to unknown documents between T1 and T2. The recognition value for unknown documents at T1 (Mdn=1) is significantly higher than the recognition value at T2 (Mdn=.66) with Z=-2.105 and p=.034 i.e. if the time between the two phases is 4 days, it has a negative impact on the recognition value for unknown documents. There is no significant result for known documents.
  Moreover, we found that the level of visibility influences this result. For shared tags (public), the average recognition value at T2 (Mdn=.66) is lower than the average recognition value at T1 (Mdn=.83), Z=-2.104, p=.033.
  Such a result is important when exploiting shared tags: the more time elapses, the more the recognition of tags decreases.

V. Conclusion

The objective of this paper was to study the effect of several human characteristics (memory, intent, knowledge level, etc.) on recollection (particularly in relation to online platforms based on tags, e.g. *Bibsonomy*[11]). We defined an experimental protocol to carry out this study. We then presented the data collected from 32 participants. Three main hypotheses were identified and verified. From these hypotheses and the results obtained we were able to determine the impact of the characteristics and the recollection measures (recall/recognition). The main results of this study are:

- Tags from expert users (high knowledge) represent a very high value-added information related to document content since these tags are chosen, in an higher proportion, outside documents. As a consequence it is really worth to exploit them;
- Age has an impact on recall (older participants have a lower recall score than younger participants).
  Moreover, age has also an impact on the number of tags chosen to describe a document (young participants use an higher number of tags than older participants);
- The recognition score at T2 (T0+7) for public tags is lower than the recognition score for of public tags at T1 (T0+3).

Unfortunately some other results cannot be accurately interpreted.
In addition, maybe the time used (4 days) between the T1 and T2 phases should be increased in order to identify the relative importance of time on recollection. An additional analysis with other participants should be envisaged to complement our results.
Moreover, there are several possible perspectives that could be added to this study. We focused on whether or not a user chooses a word inside or outside its content when tagging a document. A preliminary perspective would be to check if the salience of words in the text has an impact on the choice of tags (especially when the users knowledge is low). A second perspective would be to check the way the results presented in this paper could be integrated in a personalized system (in particular user profile management and exploitation). For instance, it could be linked to the way users choose words to formulate their queries and to help to identify query difficulty [20].

[11] https://www.bibsonomy.org/

Acknowledgment

This research is supported by the French National Research Agency (ANR) - project CONTINT ANR-12-CORD-0028.
We would gratefully thank Clara Valette and all the participants to this study for their participation.

References

[1] M. Gupta, R. Li, Z. Yin, and J. Han, "Survey on social tagging techniques," *SIGKDD Explor. Newsl.*, vol. 12, no. 1, pp. 58–72, Nov. 2010. [Online]. Available: http://doi.acm.org/10.1145/1882471.1882480
[2] B. Harry, J. William, and D. Susan, "Information behaviour that keeps found things found," *Information Research: An International Electronic Journal*, vol. 10, no. 1, 2004.
[3] G. Cabanac, M. Chevalier, C. Chrisment, and C. Julien, "Collective Annotation: Perspectives for Information Retrieval Improvement (regular paper)," in *Large-Scale Semantic Access to Content (Text, Image, Video and Sound) (RIAO), Pittsburgh, USA, 30/05/2007-01/06/2007*. http://www.le-cid.org: Centre de hautes tudes internationales d'Informatique Documentaire (C.I.D.), mai 2007, pp. 529–548.
[4] M. Chevalier, C. Chrisment, and C. Julien, "Helping people searching the web: towards an adaptive and a social system," in *International Conference WWW/Internet, Madrid, 06/10/2004-09/10/2004*, P. Asaias and N. Karmakar, Eds., octobre 2004, pp. 405–412.
[5] S. Hayman, "Folksonomies and tagging: New developments in social bookmarking," 2007.
[6] M. A. Hearst and D. K. Rosner, "Tag clouds: Data analysis tool or social signaller?" in *HICSS*. IEEE Computer Society, 2008, p. 160.
[7] Y. Hassan-Montero and V. Herrero-Solana, in *InScit2006: International Conference on Multidisciplinary Information Sciences and Technologies*, 2006.

[8] K. Knautz, S. Soubusta, and W. G. Stock, "Tag clusters as information retrieval interfaces," *CoRR*, vol. abs/1003.1048, 2010.

[9] P. Andrews, I. Zaihrayeu, and J. Pane, "A classification of semantic annotation systems," *Semant. web*, vol. 3, no. 3, pp. 223–248, Aug. 2012.

[10] A. Hotho, R. Jäschke, C. Schmitz, and G. Stumme, "Information retrieval in folksonomies: Search and ranking," in *Proceedings of the 3rd European Conference on The Semantic Web: Research and Applications*, ser. ESWC'06. Berlin, Heidelberg: Springer-Verlag, 2006, pp. 411–426.

[11] P. Lamere and E. Pampalk, "Social tags and music information retrieval," in *ISMIR 2008, 9th International Conference on Music Information Retrieval, Drexel University, Philadelphia, PA, USA, September 14-18, 2008*, J. P. Bello, E. Chew, and D. Turnbull, Eds., 2008, p. 24.

[12] M. Melenhorst, M. Grootveld, M. van Setten, and M. Veenstra, "Tag-based information retrieval of video content," in *Proceedings of the 1st International Conference on Designing Interactive User Experiences for TV and Video*, ser. UXTV '08. New York, NY, USA: ACM, 2008, pp. 31–40. [Online]. Available: http://doi.acm.org/10.1145/1453805.1453813

[13] H. H. Kim, *Intelligent Information and Database Systems: 5th Asian Conference, ACIIDS 2013, Kuala Lumpur, Malaysia, March 18-20, 2013, Proceedings, Part II*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, ch. A Semantically Enhanced Tag-Based Music Recommendation Using Emotion Ontology, pp. 119–128.

[14] S. A. Golder and B. A. Huberman, "Usage patterns of collaborative tagging systems," *J. Inf. Sci.*, vol. 32, no. 2, pp. 198–208, Apr. 2006.

[15] C. man Au Yeung, N. Gibbins, and N. Shadbolt, "A study of user profile generation from folksonomies." in *SWKM*, ser. CEUR Workshop Proceedings, P. Dolog, M. Kr?tzsch, S. Schaffert, and D. Vrandecic, Eds., vol. 356. CEUR-WS.org, 2008.

[16] P. De Meo, G. Quattrone, and D. Ursino, "A query expansion and user profile enrichment approach to improve the performance of recommender systems operating on a folksonomy," *User Modeling and User-Adapted Interaction*, vol. 20, no. 1, pp. 41–86, Feb. 2010.

[17] S. Kichou, H. Mellah, Y. Amghar, and F. Dahak, *Active Media Technology: 7th International Conference, AMT 2011, Lanzhou, China, September 7-9, 2011. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, ch. Tags Weighting Based on User Profile, pp. 206–216.

[18] T. Svensson and L.-G. Nilsson, "The relationship between recognition and cued recall in memory of enacted and nonenacted information," *Psychological Research*, vol. 51, no. 4, pp. 194–200, 1989. [Online]. Available: http://dx.doi.org/10.1007/BF00309148

[19] F. Haist, A. P. Shimamura, and L. R. Squire, "On the relationship between recall and recognition memory," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 18, pp. 691–702, 1992.

[20] S. Mizzaro and J. Mothe, "Why do you Think this Query is Difficult? A User Study on Human Query Prediction," in *ACM SIGIR Special Interest Group on Information Retrieval (SIGIR), Pisa, Italy, 17/07/2016-21/07/2016*. http://www.acm.org/: ACM, juillet 2016.