REFERENCE MANAGEMENT SOFTWARE (RMS) IN AN ACADEMIC ENVIRONMENT: A SURVEY AT A RESEARCH UNIVERSITY IN MALAYSIA

MOHAMMAD OSMANI, ROZAN MZA, BAKHTYAR ALI AHMAD, ARI SABIR ARIF

1 Department of Management, Mahabad Branch, Islamic Azad University, Mahabad, Iran
2 Faculty of Computing, Universiti Teknologi Malaysia (UTM), Johor, Malaysia
3 Faculty of Geo Information and Real Estate, Universiti Teknologi Malaysia (UTM), Johor, Malaysia
4 Faculty of Physical and Basic Education, University of Sulaimani (UOS), Sulaimani, Iraq

E-mail: m.osm2003@gmail.com, drmohdzaidi@gmail.com, bakhtyar.aliahmad@gmail.com, arisabirarif@gmail.com

ABSTRACT

Reference Management Software is used by researchers in academics to manage the bibliographic citations they encounter in their research. With these tools, scholars keep track of the scientific literature they read, and to facilitate the editing of the scientific papers they write. This study presents the results of a quantitative survey performed at a research university in Malaysia. The aims of the survey were to observe how much these softwares are used by the scientific community, to see which softwares are most known and used, and to find out the reasons and the approaches behind their usage. Manually questionnaire was distributed to the Master and PhD students at all faculties in Jun 2014. The data collected were analysed through a constant comparative analysis, and the following categories were drawn: a basic practical approach to the instrument, the heavy impact of the time factor, the force of habit in scholars, economic issues, the importance of training and literacy, and the role that the library can have in this stage. Describing the present situation, the study gives final directions to the libraries to better perform effective tasks about the matter. This study presented here is the first survey of the actual distribution and usage of Reference Management Software in a research university in Malaysia. This picture can give an important glance to Reference Management Software as one of the elements in the academic digital libraries.

Keywords: Reference Management Softwares, Citation Managers, Academic Libraries, Virtual Collaboration, RMS.

1. INTRODUCTION

This According to Telstar's definition, a Reference Management Software (RMS) enables an author to build a library of references by entering the details of each reference in a structured format. They usually support mechanisms for organizing sets of references by tagging or use of ‘folders’, and will generate references, citations or bibliographies in a range of referencing styles. Most packages support ways of importing records from library catalogues and other bibliographic data sources in order to facilitate the generation of references. In addition, many packages offer plug-ins or add-ons for Word processing software which enable authors to insert references from their ‘library’ directly into a document as they are writing. To summarize, RMS have two main functions; (1) Building a database of citations, useful for keeping track of and organize the documents useful for one's research. (2) Formatting bibliographies and citations when writing papers.

Today's packages offer more sophisticated functionalities, and their basic functions are extended through advanced features which vary from software to software. Some of them allow managing the actual full-text document together with the reference (e.g. PDF files), often including ways for annotation. Most of them take the best of the web environment providing APIs which allow integration with other software’s or other virtual environments, sharing and enriching the data, collecting them from different sources.

As technology allows seamless transmission of documents from the web to the desktop, it allows in the same way cooperation between users. In this
way a RMS can also become a virtual research environment, or a platform for a collaborator [1, 2]. Recent products inherit the features already adopted in virtual web collaboration networks, such as academic social bookmarking [3, 4].

Thinking about the role of the RMS in a large academic institution like the one described above, two questions naturally come to mind; (1) What level of awareness about RMS exists in the members of this research university? (2) What are the major trends in the usage of the RMS among the scholars?

This whole study moves directly from these two questions. The aims of the study are: (a) To explore and understand the measurements about the actual awareness and usage of RMS. (b) To understand the context in which scholars operate when dealing with citations and literature management. (c) To provide evidence-based information upon which libraries can base their strategies about services, assistance, and training.

To achieve these aims, the followings objectives are set to: (1) Verify how much users are informed about the potentials of RMS. (2) Verify whether RMS are used or not, and to what extent. (3) Establish which softwares are the most known and the most used. (4) Consider and understand the users’ behaviour in dealing with citations. (5) Ask the scholars about the reasons which influence their use.

2. LITERATURE REVIEW

To prepare the ground for a survey about Reference Manager Software at university level, we consulted all the literature about RMS and citation management in academics. We started from some of the main bibliographic databases specialized in library and information science: LISA (Library and Information Science Abstracts), and the Library Science journals published by Emerald. We performed several searches using keywords such as Reference Management Software, Citation Management, and Bibliographic Management; sometimes it was useful to use the name of software products: EndNote, Zotero, Mendeley, and RefWorks. We extended the research to more generic search tools like Google Scholar, ISI Web of Knowledge and the Mendeley Database. The database search brought up some articles; simple promotional documents or tutorial guides were not considered.

Literature about RMS focuses mostly on two main themes: on one side we find description, comparison and technical analysis of the features offered by the software packages; on the other side, we find papers about library initiatives of training and promotion. These two main threads are confirmed by McMinn [5]. Since RMS is practical tools used in real-case contexts, it is worth to look at them from the perspective of the users' behaviors and their relationship with other digital research tools. Finally, being this research about the usage of RMS, previous similar studies are reviewed.

Given this, the present review will be divided into these five areas;

2.1 Technical Aspects

The simple technical analysis of the products is very common. A good overlook on the available softwares is the Wikipedia entry “Comparison of reference management software” which is a complete and up-to date list. Specific comparative studies of features were performed by [6, 7, 8]. The column “Internet Resources” of the Public Services Quarterly 6, 2010 [9] offers a range of short reviews of different packages: Mendele [10], Zotero [11], EndNote [12], CiteULike [13], RefWorks [12]. The same approach is provided by [14, 15].

Beside these articles we also find several papers focusing on single packages [16, 17]. This is not surprising for new softwares appearing on the market: attention and analysis of the novelty is normal, and new articles are to be expected as long as old softwares develop into new products.

2.2 Virtual Collaboration

Another extremely interesting trend that is slowly gaining space in the literature is the connection between the personal citation management, its opening to the virtual collaboration and its impact on scientometrics. In a more general perspective which considers a wide range of digital tools useful to support the scholar's activity, in their article about the approach to digital libraries by researchers, Hull, Pettifer and Kell consider RMS as instruments that could enhance both personalization, social networking and collaboration, integration and accessibility [18].

2.3 Researchers Behavior

A study of RMS cannot ignore the researchers' habits and behaviour. Therefore citation management is often considered in the perspective of the whole scholar's research process [19, 20], often compared to other digital tools.

Haglund and Olsson [21] find dramatic evidences among Swedish researchers: “Their
searches seem simple, aimless, and unstructured, they do not read manuals, and they seldom use the alternative for advanced search”. The researchers understand that it is the responsibility of libraries to organize access to information, but it is not something they reflect on. Neither is it something that generates contact with the libraries with questions concerning provision of information. The researchers visit the physical library more or less frequently, but often prefer to manage on their own.

The common point rose that the usage of specific reference management tools is scarce and inconsistent. Childress considers the RMS in a practical perspective, studying them within the researchers’ needs and workflows, and reflects about the supporting role that libraries can have [22]. His article offers a wide look at the whole problem, considering both the user's behavior and the library functions, bridging the topics examined above with the second main trend found in the literature.

2.4 Role of Libraries

This second trend consists in few interesting information given about training initiatives that involved library staff [23, 24]. According to East [24], the well-established role of the library in training researchers in searching electronic databases and downloading retrieved references, leads to the involvement in this matter. East [24] points out that in many institutions the library has come to be seen as the main center of expertise in matters related to personal bibliographic softwares. The role of libraries is also confirmed by Crowley and Spencer [25] which Libraries also need to make their research management and collaboration tools such as EndNote, EndNote Web, Zotero and RefWorks easily available, and ensure that all search interfaces incorporate a straightforward citation export function [25]. Collected objective data about the distribution and the variety of approaches to the tool will be the first step to understand the actual impact of RMS in scholars' behavior. McMinn [5] explains quite convincingly the importance of such an inquiry. There are a number of reasons why it is important to examine the different approaches research libraries take in providing similar services: ensuring that the services provided are consistent with those of peer institutions; determining how services have been tailored to meet the unique needs of different institutions; determining the level of support and optimum allocation of resources [5].

2.5 Software Review

For this research, the authors decided to consider a selection of softwares, among the packages available on the market. A complete and updated list is given on the Wikipedia page “Comparison of reference management software”, which provides very detailed information about each software.

A lot of these packages are open-source projects which lack consistency in the development, and are not updated or maintained at a sufficient level. For this reason we chose to focus on the softwares which Wikipedia indicates as active and updated in 2014:

- EndNote e EndNote Web, ProCite, and Reference Manager
- Zotero
- Mendeley
- RefWorks
- Scientific Social Bookmarking: Connotea, CiteULike, BibSonomy
- BibTeX
- BibDesk and JabRef
- Papers and Bookends
- Citavi
- Qiqqa

3. METHODOLOGY

A questionnaire collected the measurable quantitative information. Manually questionnaire was distributed to the Master and PhD students at all faculties in a research university in Malaysia. The data collected were analyzed through a constant comparative analysis. The Microsoft Excel 2013 were used to compare and analyze the data. This dynamic tool can be employed in a multitude of ways, from tracking time, to finances and even sales. One of Excel’s most useful functions is the ability to develop graphs and charts from information which can then be easily analyzed.

3.1 Data Sample & Results

The sample targeted for the survey was built upon several considerations. The dimensions and the variety of the population of the research university in Malaysia, counting all Master and PhD Students from various faculties. The questionnaire was addressed to the whole scientific population. Knowing that questionnaires usually have a low rate response [17, 26]. We planned to reach a random sample of at least 10-15% of the
global population, so the data could be reasonably
generalized from the quantitative point of view.

The questionnaire collected 216 responses, reaching a response-rate of 14.2% of the initial recipients. First of all, this number confirms the general awareness about online questionnaires response rate: as Pickard warns that questionnaires produce a notoriously low response rate [26]. Second, the number of respondents may lead to the temptation of interpreting the non-responses as a lack of knowledge or interest in the topic. This should be avoided, because it would be nothing more than a conjecture not subjected to proofs. It is also very interesting to note how the survey itself raised some interest in the participants: some respondents used the open box in the end to say “This is the first time I hear about these tools”. The academic roles are equivalently divided among PhD and master students (52.3% and 47.7%).

3.1.1 Awareness and usage

The first important result is the general awareness about reference tools: only 6.5% of the respondents declare to not know and only 10.2% of them do not use any RMS software.

Figure-1 shows that EndNote proves to be the best-known software: 92.6% of respondents know or heard about it, and the 47.2% of them know about Mendeley. The other side of this data is the relatively low knowledge about alternatives to EndNote. All the other softwares seem to be mostly ignored; BibTex and EndNote Web obtain 18.5% and 17.6% respectively, and the rest are below than 10%. Data about usage show a more extreme trend. The non-usage is relevant: 10.2%, almost a 0.1 of the sample. Usages of EndNote reach more the 3/4 of the sample: barely 75.6% is the number of actual users, and just half of them (25.6%) use Mendeley. Of all the other softwares, only one is more than 5% (EndNote Web; 6.5%). It is remarkable the narrower set of softwares indicated in this answer: some softwares obtain 0 responses. Among the rich software offer and availability, scholars seem to choose a very small set of them. If we look at a correlation between the knowledge and the usage, we can obtain a percentage of “appreciation”, the percentage of those users who, knowing software, don't use a different product, in opposition to the others who know the product but don't use it. EndNote still proves to be the stronger software, and Mendeley with 25% is the next one.

The software distribution among age-ranges doesn't show any relevant result (Figure-2). There is only one fact worth of mention: the percentage of usage is higher among youger (especially for EndNote and Mendeley). Other minor notable results are the absence presence of others (BibTex, Reference Management, ProCite, BibDesk ...) among the Youngers.

Figure-2. Percentage Of Software Distribution Per Age

3.2 Reasons and Behavior

Information about user behavior and the reasons behind it have to be analyzed to be better understood. From a general point of view, we see that the most relevant reasons behind the choice of software indicate a sort of passive behavior (Figure-3). Softwares are mostly used because provided by the institution (26%) or used by the rest of the community (24%).

While the community has a strong role and gratuity and open-source collect different responses (15%), about 11% pays attention to the free of cost and only the 4% read articles about it.
From a quantitative point of view, usage of RMS varies: the number of citations saved ranges equally for less than 50 and 201-500 references (Figure-4). Obviously, more than half of the respondents have saved the number of citations from 50 to 200 (56%).

Figure-5 reveals interesting data about the general approach to the tool. The most used features are the basic ones: reference saving (29%), editing (18%) and pasting (17%) the citations when writing the paper. Fewer respondents mention organizing references for easier retrieval and management (12%), Storing pdf files of research papers (11%), and creating lists of references (10%). What impresses the most is the almost non existing usage of the RMS as a way to discover new references (3%).

3.3 Training and Support

Softwares are generally used as self-taught. About 44% of respondents declared to have followed training sessions, and 51% of respondents state that they received help by the library in using the RMS. This particular question could not be simply answered with a yes or no, so it was offered the opportunity to go deeper with the help of an open question (Figure-6).

In the Figure-7 (any Support Received by the Library), from the 110 “Yes”, only 38 provided details. Table-1 shows that 26 of them generally refer to the EndNote when libraries provided copies of the softwares for their members together with information, support, and training sessions, and the rest (12) attended a short course on how to use Mendeley software provided by the faculties. Of the 106 “No”, 27 provided details. Most respondent admit that they just “never asked”, or “never heard about any initiatives”. This case reveals how much scholars may lack of initiative or time to dedicate to the subject, but also how weakly perceived is the
role of the library. Some don’t consider the library as a potential support in the matter: “I heard about these tools from other colleagues; I recommend it to advertise all RMS to be known for others, especially in field of Engineering; Practice and self-teaching are better than any courses”. In at least five cases, the respondents consider themselves already skilled enough: “It is easy to use, and no need to attend workshop”.

Table 1. Details of any Support Received by the Library

<table>
<thead>
<tr>
<th>Yes (110)</th>
<th>No (106)</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Libraries provided copies of the EndNote softwares with information, support, and training sessions.</td>
<td>- Never asked or heard about any initiatives.</td>
</tr>
<tr>
<td>- Attended a short course on how to use Mendeley software provided by the faculties.</td>
<td>- It is easy to use, and no need to attend workshop.</td>
</tr>
<tr>
<td>- Heard about these tools from other colleagues.</td>
<td>- Heard about these tools from other colleagues.</td>
</tr>
</tbody>
</table>

When asked if they ever suggested the tool to others (Figure 8), the majority replied “Yes” (81% against 19%). A minor percentage of the people who follow a research project, even on a master thesis level, underline the importance of managing such a tool by suggesting its usage to others.

The choice itself of the software is never problematic: software is chosen because already used or suggested by other colleagues or because it’s dominant in the community. The technological context is also a key factor: according to the operating systems and word processor used, the most compliant software is adopted. It is interesting in this matter to note how EndNote was often already used: that initiative was suggested and promoted by professors who already knew the product, and asked the university to make it available. Anyway, this enhanced the presence of a legal copy of the product in every lab, so the new researchers in need for a RMS often found it already available at hand.

4. DATA ANALYSIS

The first objectives were actually achieved by the questionnaire: to verify how much users are informed about the potentials of RMS; to verify whether RMS are used or not, and to what extent; to establish which softwares are the most known and the most used.

Awareness is relatively high in terms of quantity (i.e., the majority of people know about RMS) but low in terms of quality: very few are the known softwares, and low is the relationship between knowledge and usage. RMS is used by the 83% of users. The questionnaire clearly declares EndNote and Mendeley as the most used softwares, and a very low range of alternatives: BibTeX, Papers, and Reference Manager, all of them with incomparable low numbers. For these initial goals, which served the aim of measuring the distribution of RMS, the numbers basically speak for themselves. The remaining objectives were: to understand users’ behavior; to find out the reasons behind the use. They were achieved through the questionnaire, which also helped to understand the reasons and put them in the context. The following categories, or concepts, can be drawn from the data presented above:

4.1 Basic Practical Approach

For the researchers involved in this survey, RMS are tools, and nothing more. They are used when needed (when writing a paper which requires a reasonable number of references), they are used in their basic functions, and they need to work fine.
They are a tool do a job better, so they are not used for the sake of using them, or for the pleasure, or for curiosity. This explains the approach expressed in the questionnaire, which shows a very basic need underlining its usage. The small set of most used softwares is a clue for this, too: at this level of mastery of the tool, all the softwares look all the same; therefore scholars stick to what is already well known and tested.

This also emerges about the technological issue: participants in the survey don’t show interest in the technological implications of the tool, as long as it works. This leads to be closed against additional extended features, or to paradigm changes: the ignorance about the world of virtual science and networking collaboration explains how little today scholars are aware of the opportunities provided to scientists by the web environment.

4.2 Time Factor

Time is a crucial factor in everything. This is strictly correlated to the previous category. No scholar is willing to spend his time on something different than his work. Experimentation and curiosity are used for the scientific activity, not for writing its results. Therefore, everything additional in the process must speed their work and save time, not the opposite. This applies to all the aspects: choice of a software and discovery, deep knowledge of its functionalities, training and learning sessions.

It is worth noticing that citation management is something often deeply rooted in the research process: yet it is often perceived as an element of minor importance, ready to be sacrificed towards other needs (such as looking for findings). Overwhelmed with more urgent needs, the refinement of technological skills in the research process affects citation management in a very small part. It is also true; on the other hand, a more proper training on RMS could help saving time.

4.3 Habit

A general laziness, or force of habit, prevents change. Softwares are used for a long time before they get changed to new, better performing, solutions. Even when researchers feel a push for change, or they feel unsatisfied with their current product, the issue is postponed. This attitude prevents scholars to discover new products or new features. When a RMS is used, generally it's because a former experience by some colleagues proves it useful. If the tool is suggested, then a scholar begins to test it and use it; if not, it is very unlikely that someone is willing to experiment something new on his own. When this happens, it generally leads to frustrating and unsuccessful experiences.

This is shown by the fact that a very low range of softwares is actually used, compared the softwares known. Finally, the fact that the University acquired and distributed licenses of EndNote made the faculties stick with this software without worrying about other alternatives. Now that the licenses are not purchased anymore, it will be interesting to see how scholars will change their approach.

4.4 Economic issues

Economic issues are always important, even when selecting software. Especially in these last years when findings are being cut year by year, scholars are careful about the way money is spent. Everything that can save money is welcome: this applies to softwares as well. The cheaper solution is preferred. Yet this seems true more on the intentions than in the practice: only 13% of the participants in the questionnaire actually indicate it as a reason of choice. The habit of already-in-use tools is stronger than the need to move on better instruments. Often the economic constraint is not strong enough to push people to experiment alternatives.

4.5 Training and literacy

Some respondents recognize that they need more information about RMS. Others seem to be confident about their current knowledge, but then they reveal how many useful opportunities they don't know. If we compare the answer to the questionnaire, which says that 49% never received or asked any support, which show how basic is the general knowledge of the tools and their functionalities, it is clear how impact has the lack of specific training. Even if not stated explicitly, there is need for training and literacy. Results clearly show how low the awareness is because scholars don't know RMS at all and don't have time to go deeper and improve their skills beyond the self-taught basics.

There are no common practices in the training to RMS: even if everyone's story can look the same (I use it because suggested by a colleague” or “I use it because everybody else in the lab uses it) every scholar has his own path to it. The usage of a RMS is more part of a “tacit knowledge” present in the research environment, rather than a conscious part of the set of skills and methods of a researcher. It is remarkable how every concept examined so far – shallow knowledge, time constraints, and economic
awareness – can be considered within a set of aimed training initiatives.

Given this, a strong condition arises: any kind of training must be tailored to the actual needs. This is heavily connected to the above concepts “Basic practical approach” and to the “Time factor”. If RMS serves the purpose of facilitating the research process and saving time, any training on it must not go in the opposite direction.

4.6 Library Role

From the data collected, it is clear that librarians, as information experts, must have a more active role in RMS support. But this role must consider in the more general context of the library impact in a community. The survey shows that library staff skills are mostly not perceived, therefore scholars are alone when they face reference management issues. This creates a separation between the library and the academics instead of bringing a mutual dependence. If the library assumes the role of information assistants and technology experts, it can introduce their members to better solutions, improving their workflow and saving them the burden of testing unknown products; it can propose and support standards of use; it can provide training so that the knowledge and skills are equal; it can inform about updates or new solutions. It can be the link between the world of technological information solutions – such as RMS – and researchers' needs.

5. CONCLUSION

From the data discussed above, it is possible to draw some final considerations, trying to connect the data results with the aims of the research. RMS is considerably used across scientists, even though on a simple level. There is no unique approach to a tool such as a RMS; they are used mostly as a personal tool, so its usage is tailored to the single person. They are nothing more than tools to facilitate the preparation of a final publication. RMS is used in their basic functions, without much consideration of the interesting feature. In the same way, virtual collaboration is a concept and praxis still far from the habits of the University. Collaboration itself is very important, essential part of the research job some say, but technology is not perceived as something helpful with that. Recently, web and social oriented software like Zotero or Mendeley stepped on to the stage providing new ways to manage scientific literature.

There is a lot of room for the library to be active in this process. Responses let emerge needs such as: information, training, and guidance. Library is not the keeper of resources anymore, but also the keeper of bibliographic tools. RMS require a lot of time and skills that researcher seldom have; a professional expert in these tools could help the scholars guiding them across the wide range of packages, across the basic functions, focusing on problem-solving activities. This could be an extremely cost-benefit effective initiative. This consideration confirms what is said in the literature. East already noted the relationship between bibliographic support and reference management training. He recognizes the well-established role of the library in training researchers in searching electronic databases and downloading retrieved references. From here it was only a short step to beginning to train researchers in the management of those references [24]. Every library, though, exists in a specific context, which is not the same everywhere. East [24] notes that in many institutions the library has come to be seen as the main Centre of expertise in matters related to personal bibliographic software. This has not happened completely yet at this research university, but the survey suggests that it should, and that a loud call for a new commitment is given.

This could apply also to students, as a part of information literacy strategy, but the survey doesn’t provide clear evidence on how important is for students to receive this sort of training. This confirms what was already noted by Duong [27], Since most undergraduate science courses do not require writing-intensive research papers, many departments have not seen the relevance of IL instruction.

6. FURTHER STUDIES

The present research shows important limits, both in the methodology as in the findings. Methodological limits due to practical constraints were predicted at the beginning of the studies. This survey, the first of this type at least in Malaysia, gives an important picture of the distribution and usage of RMS in a big academic environment. The findings shown above give room for more areas of inquiry.

It would be extremely interesting to perform a similar survey to other universities, and compare the results finding patterns, similarities and differences among them to reach a nationwide overview of the phenomenon.

It would be useful to perform a similar survey which includes the HSS disciplines, and give scientific evidence to the original assumption that these academic fields are unaware of RMS.
Finally, it should be worth to analyze the results from the user context point of view, considering factors such as age more deeply than the present research. Many of the concepts identified and discussed above can gain interest if looked considering the age factor: technological issues such as networking, open-source philosophy, openness to novelty and interest in training can affect very differently people of different age. Also, having proved that habit is a strong factor, searching for patterns of behavior among different age ranges could lead to important understanding on how the phenomenon is likely to change in the next future.

References:


