

CHAPTER 4

EXPERIMENTAL RESULTS

This chapter presents the results evaluation by using both qualitative and quantitative approach.

This chapter is organized as follows. Section 1 presents results evaluation by using qualitative approach. Section 2 presents results evaluation by using quantitative approach. Section 3 presents results discussion. Section 4 presents application of thesis. And section 5 concludes the chapter.

4.1 Results Evaluation by Using Qualitative Approach

This section presents the table and graph of satisfaction from three group of users derived from questionnaires. Domain expert group composed of two experts in agriculture field of Naresuan University. Expert librarian group composed of four expert librarians from Naresuan University and Nakhon Sawan Rajabhat University. And general user group composed of 500 students of Nakhon Sawan Rajabhat University. The questionnaire is used in the result evaluation have 4 topics composed of correctness of results, easy usage of the prototype, saving time in metadata extraction/searching and overall satisfaction of the prototype system. In questioners,

each topic contains score from 1 to 10 points. The summary of system evaluation by using qualitative approach is shown in table 4.1.

Table 4.1 Results of satisfaction.

Scenario	Domain experts	Expert librarians	General users
Correctness of Results	10	8	9
Easy Usage of Prototype	8	8.5	8.34
Saving Time in Metadata Extraction/Searching	8	10	8.5
Overall Satisfaction	8	8	9.5
Average	8.5	8.625	8.835

From table 4.1, the average satisfaction of domain experts, expert librarians and general users are 85%, 86.25% and 88.35% respectively.

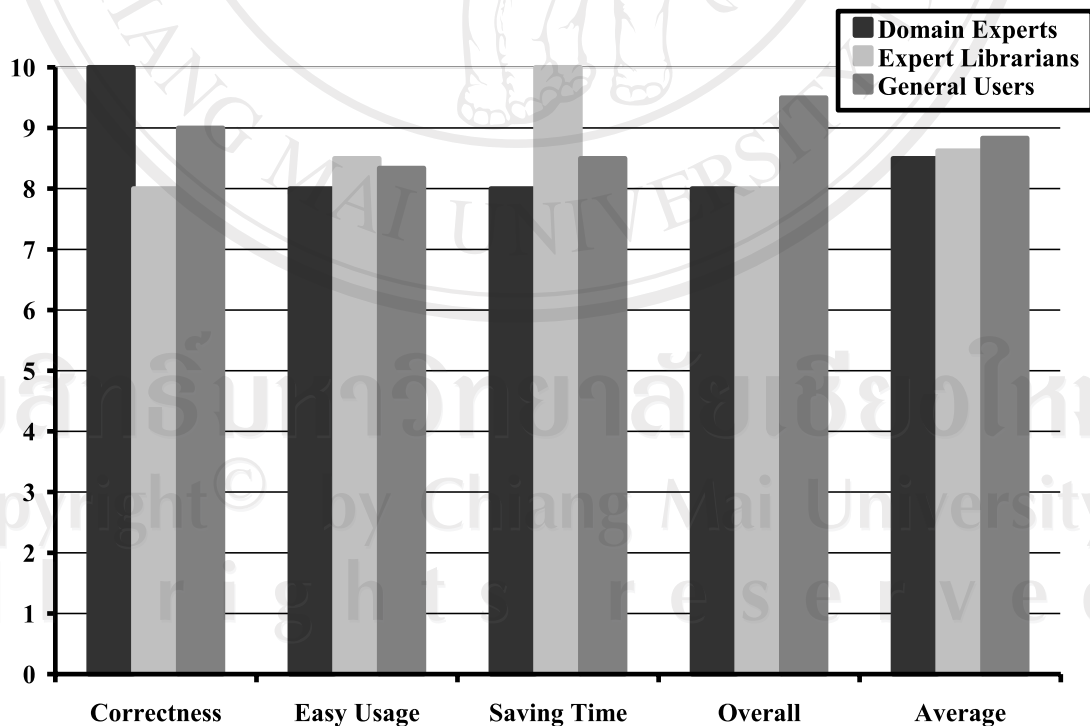


Figure 4.1 Comparison of satisfaction in four scenarios.

Figure 4.1 shows the comparison graph of satisfaction of user classified by the evaluated group composes of correctness, easy usage, saving time and overall. The group of domain expert gives more correctness score than other group. The group of expert librarian gives more saving time score than other group. The group of users gives more average score than other group.

4.2 Results Evaluation by Using Quantitative Approach

This section discusses results evaluation by using quantitative approach. The sample of experiment based on two thousand problem cases given by the domain experts. The prototype is evaluated based on the following three widely used methods (Baeza-Yates and Ribeiro-Neto 2002) precision, recall and F-measure to investigate performance of the web-based KMS.

In quantitative evaluation for the research, four tables are considered for the following scenario: (1) the results of extraction by only expert librarian (see table 4.2); (2) the results of extraction by using prototype system with 0 initial revise case without expert guidance (proposed solution) (see table 4.3); (3) the results of extraction by using prototype system with 0 initial revise case with expert guidance (confirmed solution) (see table 4.4); (4) the results of extraction by using prototype system with 61 initial revise cases with guiding from expert (full efficiency confirmed solution) (see table 4.5). For observing the characteristic of prototype system, we collect 2,000 Thai theses which content related to sufficient economic and Thai folk wisdom to make the data set then dividing into 10 groups. Each group contains

randomly selected 200 from 2,000 theses which are D1 to D10. Finally, a comparison of scenario 1 to scenario 4 and a comparison of average precision, recall and F-measure are presented in form of graphs.

Table 4.2 Results from expert librarian.

Data set	The results of extraction by expert librarian		
	Precision	Recall	F-measure
D.1	0.6723	0.7175	0.6942
D.2	0.6623	0.7529	0.7047
D.3	0.7217	0.7927	0.7555
D.4	0.7027	0.7640	0.7321
D.5	0.6713	0.7779	0.7207
D.6	0.6600	0.7486	0.7015
D.7	0.6687	0.7397	0.7024
D.8	0.6553	0.6850	0.6698
D.9	0.6300	0.6779	0.6531
D.10	0.6953	0.7624	0.7273
Average	0.6740	0.7419	0.7061

From table 4.2, the maximum precision, recall and F-measure are 72.17%, 79.27% and 75.55% respectively. The minimum precision, recall and F-measure are 63.00%, 67.79% and 65.31% respectively. The average precision, recall and F-measure are 67.40%, 74.19% and 70.61% respectively.

Table 4.3 Results from prototype system without guiding from expert.

Data set	The results of extraction by using prototype system with 0 initial revise case without guiding from expert		
	Precision	Recall	F-measure
D.1	0.6393	0.6823	0.6601
D.2	0.6207	0.7056	0.6604
D.3	0.6597	0.7246	0.6906
D.4	0.7447	0.8097	0.7758
D.5	0.7277	0.8432	0.7812
D.6	0.6070	0.6885	0.6452
D.7	0.6960	0.7699	0.7311
D.8	0.7583	0.7927	0.7751
D.9	0.7273	0.7826	0.7540
D.10	0.6483	0.7109	0.6782
Average	0.6829	0.7510	0.7152

From table 4.3, the maximum precision, recall and F-measure are 75.83%, 84.32% and 78.12% respectively. The minimum precision, recall and F-measure are 60.70%, 68.23% and 64.52% respectively. The average precision, recall and F-measure are 68.29%, 75.10% and 71.52% respectively.

Table 4.4 Results from prototype system with 0 retrieved case.

Data set	The results of extraction by using prototype system with 0 initial retrieve case			
	Precision	Recall	F- measure	Case revise
D.1	0.8117	0.8662	0.8381	64
D.2	0.7547	0.8579	0.8030	52
D.3	0.8003	0.8792	0.8379	61
D.4	0.7920	0.8612	0.8251	62
D.5	0.7437	0.8617	0.7984	55
D.6	0.7457	0.8457	0.7926	60
D.7	0.7643	0.8455	0.8029	58
D.8	0.8110	0.8477	0.8290	76
D.9	0.7997	0.8605	0.8290	71
D.10	0.7880	0.8640	0.8243	72
Average	0.7811	0.8590	0.8180	63.1

From table 4.4, the maximum precision, recall and F-measure are 81.17%, 87.92% and 83.81% respectively. The minimum precision, recall and F-measure are 74.37%, 84.55% and 79.26% respectively. The average precision, recall and F-measure are 78.11%, 85.90% and 81.80% respectively. The number of average revise case is 63.1 cases.

Table 4.5 Results from prototype system with 63 retrieved cases.

Data set	The results of extraction by using prototype system With 63 initial retrieve case			
	Precision	Recall	F- measure	Case revise
D.1	0.8790	0.9381	0.9076	5
D.2	0.8750	0.9947	0.9310	7
D.3	0.8843	0.9714	0.9258	8
D.4	0.8720	0.9482	0.9085	7
D.5	0.8617	0.9985	0.9250	5
D.6	0.8793	0.9974	0.9346	6
D.7	0.8513	0.9417	0.8943	4
D.8	0.9113	0.9526	0.9315	10
D.9	0.8987	0.9670	0.9316	4
D.10	0.8917	0.9777	0.9327	8
Average	0.8804	0.9687	0.9223	6.4

From table 4.5, the maximum precision, recall and F-measure are 91.13%, 99.85% and 93.46% respectively. The minimum precision, recall and F-measure are 85.13%, 93.81% and 89.43% respectively. The average precision, recall and F-measure are 88.04%, 96.87% and 92.23% respectively. The number of average revise case is 6.4 cases.

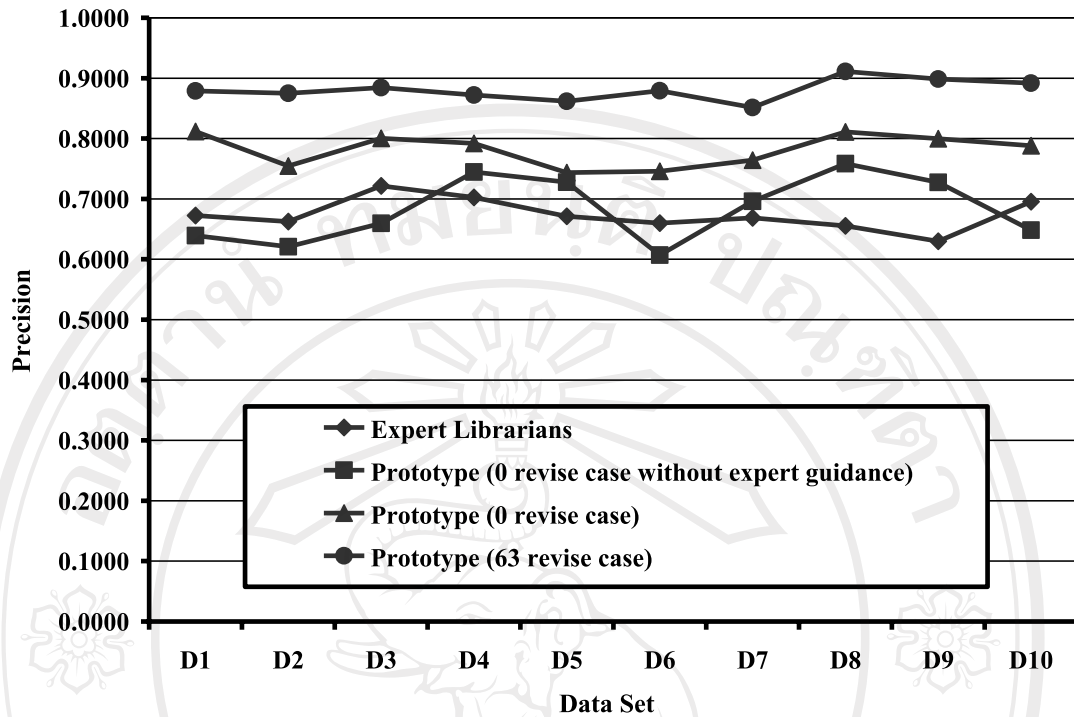


Figure 4.2 Comparison of precision in four scenarios.

From figure 4.2, the best result is the results of prototype system with 63 revise cases. The secondary is results of prototype system with 0 revise cases with expert guidance. And the result of prototype system with 0 revise cases without expert guidance and the result of expert librarians are as the same level.

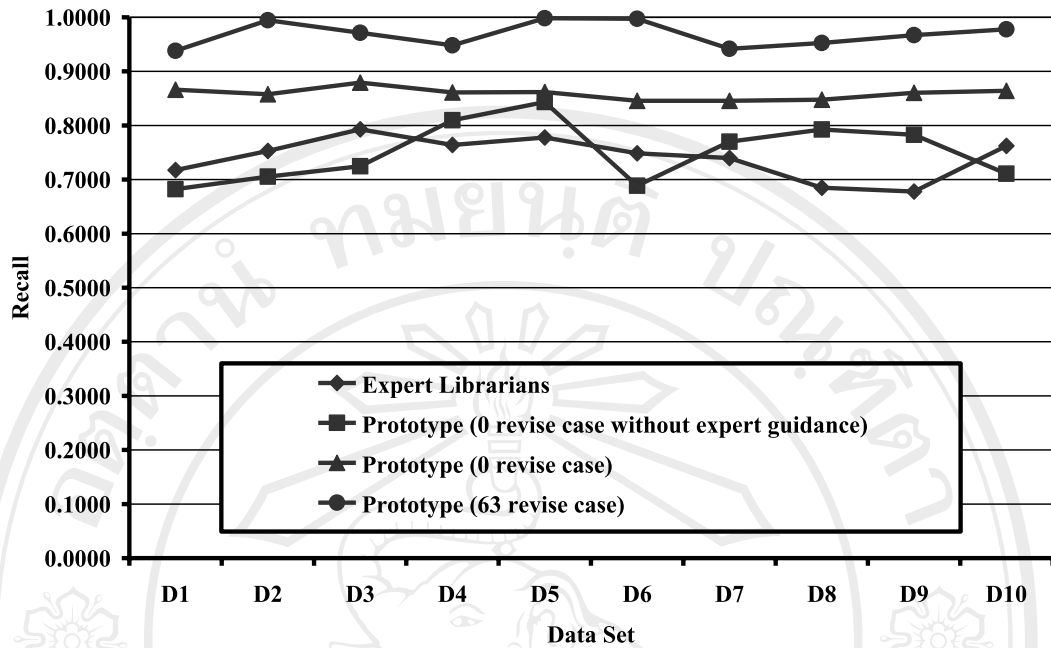


Figure 4.3 Comparison of recall in four scenarios.

From figure 4.3, the best result is the results of prototype system with 63 revise cases. The secondary is results of prototype system with 0 revise cases with expert guidance. And the result of prototype system with 0 revise cases without expert guidance and the result of expert librarians are as the same level.

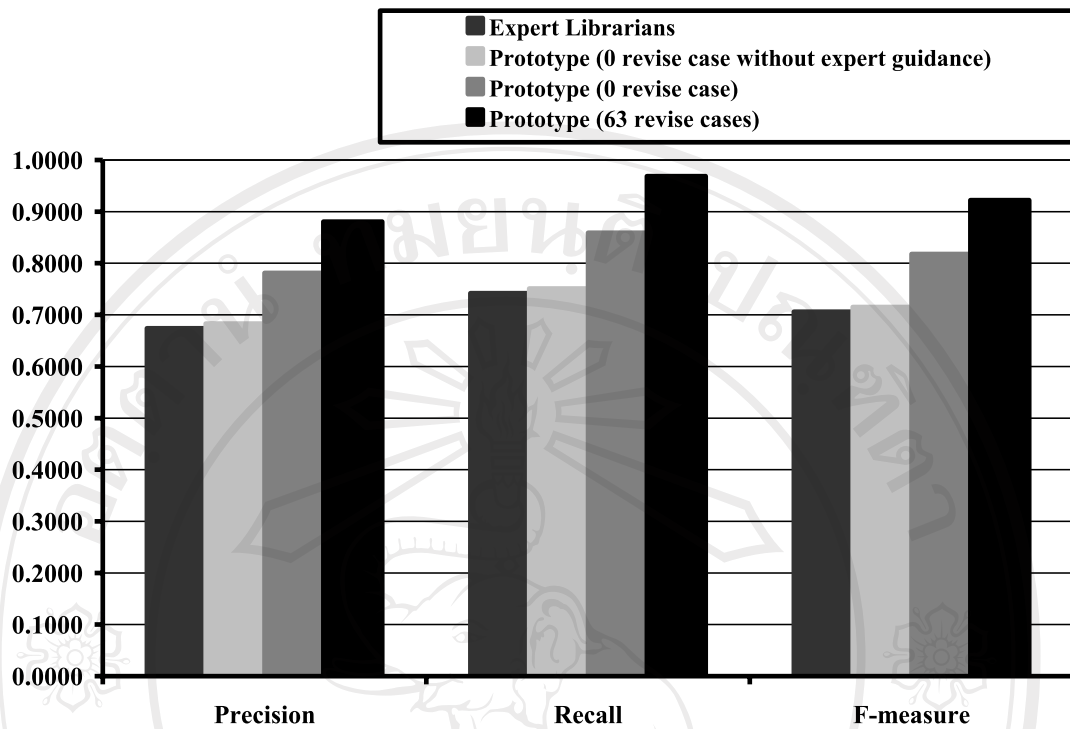


Figure 4.4 Comparison of correctness in three scenarios.

Figure 4.4 shows the comparison of Precision, Recall and F-measure of four result scenarios.

4.3 Results Discussion

This section presents the discussion of using Dublin Core Metadata Element Set, Case-based Reasoning, Information Extraction and Knowledge Management System with Thai language.

Regarding to apply Dublin Core Metadata Element Set with Thai language, since most Thai documents used in the research do not prepare the details essential for the

development of Dublin Core metadata, the data in each field of metadata is therefore incomplete as the designed Dublin core template. The number of field in the designed Dublin Core template used for the precision value evaluation equals to number of Thai documents multiply by 15 (number of standard Dublin Core metadata) or 30,000 of the total fields of the Dublin Core template. The designed Dublin Core template use for calculating recall value for evaluation results from the expert librarian who analyzes the number of real metadata field to be used for calculating the recall value. The calculated part of F-measure is to calculate the average number between precision and recall.

Among these measures, precision is a measure that directly indicates metadata extraction performance. It shows the performance by the ratio of the number of correct metadata suggested by the system to the number of ideal answer. However, recall and F-measure do not exactly express metadata extraction performance but they are useful for indirectly measuring the performance. Recall represents the ratio of the number of correct extracted metadata suggested by the system to the number of actual answer. F-measure represents the compromise between precision and recall. Recall and F-measure are one of the most important performance measures in information retrieval system.

The result of applying CBR with Thai language indicates that CBR technique can be effectively applied with Thai content. Since the existing research in CBR with Natural Language Processing is available in English language only. This research discover that CBR techniques can be effectively applied with Thai language then it

can be applied with some Asian language which grammar similar to Thai language such as Chinese, Japanese and Korean.

Regarding to apply Information Extraction theory with Thai language, since the architecture used in the research is a standard architecture for research in Natural Language Processing which composed of four module: Tokenization, Lexical processing, Syntactic processing and Domain analysis. Then, the implementation Information Extraction in Thai content does not cause problem. But the techniques used in each module of Information Extraction architecture must be change to appropriate with those languages. The important factor of applying Information Extraction theory with Thai language and some Asian language is that the selected techniques in each module must be suitable for the type of work.

In relevant to apply Knowledge Management System with Thai content, since there is not Knowledge Management System that support collaboration between Thai domain expert and librarian. This research is in the first step of development of framework and architecture of Knowledge Management System for knowledge sharing and reuse between Thai domain expert and librarian. The idea of the research can be applied to develop the Knowledge Management system that supports collaboration between Thai domain knowledge and Thai worker.

4.4 Application

This section shows how the research is utilized by comparing with one of the most profound library search systems which is used in most of the libraries in Thailand namely “Online Public Access Catalog (OPAC)”. The main objective is to compare the method of storing document specifications of Dublin core metadata with the data component stored in the OPAC system.

In OPAC, the data description composes of Author, Title, Edition, ISBN, Call Number, Subject, Location and Status. The 500 students from the Faculty of Agriculture, Nakhon Sawan Rajabhat University are selected to test the document search of both systems. Each student is capable of searching only one document specification. The maximum of 5 times to change the searching keyword are allowed so that the result will be found or unfound. Figure 4.5 shows the process as describe above. All the document search results are collected from 500 students and fill in the comparison table as shown in table 4.6.

Table 4.6 Comparison of Prototype and OPAC.

Round	Prototype (Dublin Core)			OPAC		
	Hit Rate	Hit/Round	Total Hit	Hit Rate	Hit/Round	Total Hit
1	456	91.20	91.2	348	69.60	69.6
2	23	52.27	4.6	47	30.92	9.4
3	11	52.38	0.8	38	36.19	7.6
4	4	40.00	2.2	16	23.88	3.2
5	6	100.00	1.2	9	17.65	1.8
Total	500	100.00	100	458	91.60	91.6
Unfound	0	-	0	42	-	8.4

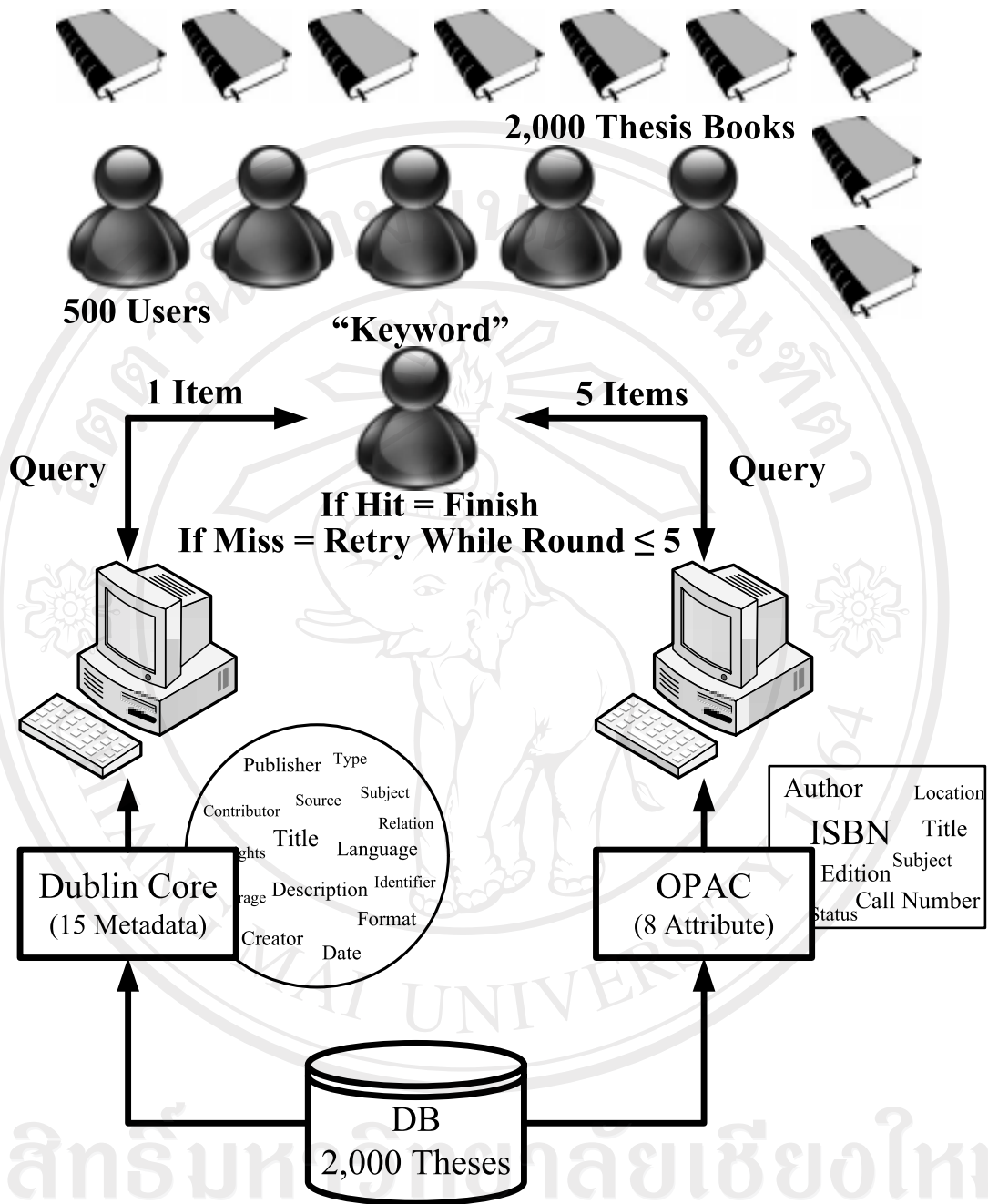


Figure 4.5 Dublin Core vs. OPAC.

Table 4.6 shows that the correctness percentage of finding the document at the first time search by Dublin Core metadata approach is greater than that the OPAC approach by 21.60%. Moreover, the correctness percentage of finding the document

within the 2nd to 5th time search of Dublin Core metadata approach are greater than that the OPAC approach by 21.35%, 16.19%, 16.12% and 82.35% respectively. The unfound document by Dublin core metadata search is also less than the OPAC search by 8.40%.

4.5 Conclusion

The chapter presents the evaluation results of the prototype. It can be seen that the prototype was able to adapt new knowledge using the knowledge adaptation process of CBR cycle. This way, old cases can be used to solve new problems and new cases can be adapted as new knowledge in the knowledge repository to solve future relevant cases.