中正嶺學報 第四十七卷 第二期 民國 107.11
JOURNAL OF C.C.I.T., VOL.47, NO.2, NOV, 2018

K2FCQE: A Hybrid Query Expansion for News Retrieval

Liang-Chu Chen*, Wen-Tsan Chao, and Chieh-Pin Lin

Department of Information Management, National Management College, National Defense University

ABSTRACT

Most search engines retrieve documents strictly based on keywords because of which other content that is similar in idea is not easily accessible. Therefore, query expansion becomes imperative, for which ontology is a critical foundation. There are two purposes of this research. The first is to design a novel Keyword to Formal Concept Query Expansion (K2FCQE) algorithm to automatically construct the relationship between ontology and vocabulary concepts and then to proceed to query mode verification. The second is to develop a prototype of a Military News Retrieval System based on the K2FCQE method (K2FCQE-MNRS). The results of this research verify that K2FCQE is more efficient than other query expansions methods and that the K2FCQE-MNRS is helpful for users to search military related news.

Keywords: ontology, formal concept analysis, query expansion, K2FCQE

K2FCQE：一套支援新聞檢索的整合式查詢擴展服務

陳良駒*　趙文璨　林杰彬

國防大學管理學院資訊管理學系

摘　　要

現有搜尋引擎多以關鍵字為基礎進行查詢，但卻不易獲得與檢索詞概念相關的內容，故具語意關係的查詢擴展機制更顯重要。相關方法已應用於資訊管理、圖書館學與醫學等領域，目前仍少見於軍事範疇之運用。本研究目的如下：(一)設計一個詞語-正規概念查詢擴展(K2FCQE)演算流程，以自動化方式建立領域知識本體及語彙概念關係，並同時進行模式驗證；(二)開發以 K2FCQE 為基礎的軍事新聞檢索系統之雛型。研究成果驗證 K2FCQE 檢索效能優於其他查詢擴展方法，同時軍事新聞檢索系統亦能協助使用者擴展相關的查詢概念。

關鍵詞：知識本體，正規概念分析，查詢擴展，K2FCQE

文稿收件日期 106.10.18; 文稿修正後接受日期 107.5.4;*通訊作者
Manuscript received October 18, 2017; revised May 4, 2018;* Corresponding author

I. INTRODUCTION

The primary purpose of information retrieval is to enable searches to meet the user's knowledge needs. However, the rapid development of information technologies and the Internet has resulted in the accumulation of extensive amounts of documents and information resources that make information search and retrieval very difficult. In this era of knowledge economy, people are changing their needs from a "demand for information" to a "thirst for knowledge." Simple information retrieval mechanisms are thus becoming increasingly inadequate to meet the needs of most users.

Today, most search engines have limited capacity for information retrieval because they are restricted to key words and cannot obtain other content related to the query terms. The main purpose of query expansion is to effectively improve information retrieval and provide more effective and relevant results [1] by obtaining concepts and terms related to the original query through various technologies and methods.

The costs of constructing information retrieval systems for specific areas of information can be reduced by text mining, conceptual graph extraction, ontology, and other techniques [2]. Therefore, how these technologies can be used for constructing and expanding the domain knowledge is an important developmental direction in of the context of current knowledge management tools. Here, ontology is an important foundation that can support the expansion of the semantic search. Although ontology is a hierarchical structure used to describe a field of knowledge [3] and has been deployed in information retrieval [4], information classification, image retrieval, cross-language retrieval [5], library science, and medicine, it remains new to military usage.

This research aims to formalize concept analysis based on the aggregation of military terms and military reports to automate the design process of constructing a military ontology. We reveal a hybrid method called Keyword to Formal Concept Query Expansion (K2FCQE) [6] and compare its performance with other query expansion methods. We then use the Military News Agency (MINA, http://mna.gpwb.gov.tw/) website as an experimental object to develop prototypes of K2FCQE-based Military News Retrieval System (K2FCQE-MNRS), which can help users expand their query terms.

II. QUERY EXPANSION

According to the research of Spink et al. [7], when searching for information on the web, 32.5 percent of users revise their original inquiry and 29.3 percent add more than one keyword to re-search. As one single word or phrase can cover thousands of keywords, using one keyword may not be enough to convey the concept the user intends to search. Query expansion, originally proposed by Rocchio [8], is based on the concept of finding synonyms or semantic-related terms for the search word and adding them to the original query to expand the search query [9]. It can effectively improve the recall of information retrieval system [10]. There are different methods of query expansion that can be categorized into manual, interactive, automatic, and hybrid modes [9, 11]; in terms of an expanded source, it can be divided into global expansion and local expansion [12]; in terms of knowledge models, thesauri and ontologies are both frequently used [13].

2.1 Ontology-Based Query Expansion: Formal Concept Analysis Method

2.1.1 Formal concept analysis

Formal Concept Analysis (FCA) is a data analysis theory that can construct conceptual structures from existing datasets [14]. It was first proposed by Wille [15] and is now being widely applied in various fields, including medicine, psychology, sociology, and information science [16, 17]. According to the definition from Ganter and Wille [18], the mathematical basis of FCA is the Lattice Theory. It indicates that a formal context C is composed of G, M, and I, where G is the collection of Objects, M represents the attributes, and I is the binary relations of G and M. The formula is shown as "$C = (G, M, I)$," where $I \subseteq G \times M$, meaning that object g and attribute m have the characteristics of (g, m) ∈ I.

Although both FCA and domain ontologies aim at modeling concepts [16], FCA provides a rigorous framework for defining and structuring concepts and is critical for developing domain ontologies [19].

2.1.2 Ontology-based query expansion

In the field of knowledge management, ontology is not only a method to describe the presentation of domain knowledge but also a mediator to present knowledge on a computer. The definition of ontology is thus as Gruber [20] formulated: an explicit specification of a conceptualization. Van Heijst et al. [21] thought of ontology as the presentation of the relations between relevant concepts using concrete knowledge orders that are subject to specific fields and tasks. Corcho et al. [22] suggested that the substance of ontology includes simplified modes (such as conceptual hierarchical relationships) and overall modes (such as details and restrictions of knowledge). Further, Noy and McGuinness [23] considered ontology as a definite description of common concepts in a specific field, including the features, attributes, and limitations of the concept, to form a shared domain knowledge base. Therefore, the purpose of ontology is to use systematic methods and procedures to define concept elements, including genre, attributes, and relations, and then transform the tacit concept into explicit structures that can be presented in common vocabulary so that users with or without the related background can access and share the domain knowledge.

Some scholars also support the query expansion method with an ontology perspective. Voorhees [24] integrated ontology and query expansion techniques to propose the ontology-based query expansion research; Bhogal et al. [25] compiled ontology-based query expansion research and classified it into common ontology and special area ontology based on the applications. For constructing ontology automatically, Boonchom and Soonthrnphisaj [26] proposed an algorithm, the Automatic Thai Legal Ontology Building (ATOB), to generate and expand ontology using Thai legal terminology. Conesa et al. [5] proposed a method to improve traditional web search that involves dividing the search term, integrating the ontological advantages of ResearchCyc and WordNet to analyze the semantic concept of the search term, and then adding the integrated result to the original query to expand the search. Yokoyama [27] analyzed the query terms and then combined them with Japanese WordNet to obtain the synonym and hyponym for query expansion. Thus, all research to date indicates that ontology-based query expansion can extract new terms from the knowledge represented in the classes, properties, and relationships of an ontology [13] and can effectively improve retrieval performance.

2.1.3 FCA-based query expansion method

Information retrieval services often use ontology to support the measurement of similar concepts, and several scholars are currently using FCA to propose a new similarity measurement model [16, 28-31]. Formica [16], for example, applied FCA and a conceptual grid to improve the current analysis of ontology while assessing and evaluating the similarity of concepts. On the other hand, Weng et al. [30] proposed using FCA-based query expansion method to first construct the conceptual hierarchy of ontologies and then use the binary relation matrix formed by news and vocabulary to express the independent, interrelated, and inheritance relationships that form the conceptual diagram of ontology.

When the concepts of FCA are extensive, it takes longer to calculate concept similarity. Alqadah and Bhatnagar [28] proposed a weighted FCA-based similarity measurement to better calculate the similarity of concepts through objects and attributes. Given that A1 and A2 represent the object elements along the object collection while B1 and B2 represent the attribute elements of attribute collection, every concept is composed of objects and attributes (e.g., C1 = {A1, B1}; C2 = {A2, B2}). The concept-based similarity measurement is as follows:

$$\mathrm{S}(C_1, C_2) = w * S(A_1, A_2) + (1-w) S(B_1, B_2) \quad (1)$$

where $0 \leq \omega \leq 1$ and S is any metric similarity measures (e.g., Jaccard index).

However, the shortcoming of the FCA-based similarity measurement is that when the user keys in a query word that does not exist in the ontology, the concept will not match any file, and thus there can be no recommendation through query expansion.

2.2 Keyword-Based Query Expansion: Keyword to Concept Method

When the hierarchy of ontology concepts is used to search and recommend files, the search term and its relationship to ontology concepts must be known. Tian et al. [32] presented a query expansion model with global expansion features called K2CM (Keyword to Concept Method). K2CM is composed of two parts: the first is the query word, file, and relationship between the concepts, and the second part is the local co-occurrence degree of the query word and concept based on an effective context.

The first part of K2CM consists of calculating the relationship between the query word and concept word by analyzing the attributes of the query word, file, and concept. For example, the file and concept can form a binary relation matrix through the relevant technology, and the user's query word may exist in numerous files where each file belongs to one or more concepts, thus describing a subordinate relation between the file and concept. Meanwhile, there is a subordinate relation between the query word and concept through the file. In this way, the correlative degree of the query word to a concept can be determined by calculating all the corresponding concepts of the files found from the query word. The latter part of K2CM refers to the research of Lu and Bai [33] who treated all the files as a large text file and took the effective context for the range of common word assessment; that is, they treated the query word as the core, searched all the files for the query word, and captured and calculated the co-occurrence word from the left eight words to the right nine words of the query word.

As these two parts are important factors that affect the correlation, the interaction can more effectively illustrate the correlation degree of the query word and concept. Therefore, the K2CM calculates the conceptual similarity by multiplying the two factors to present the correlation between the query word and concept. The advantage of K2CM is that if the query word does not exist in the ontology, the concept word expansion can still be conducted.

III. RESEARCH DESIGN

3.1 Research Framework

The research framework of this study can be divided into four development stages, namely data acquisition & preprocessing, ontology building, query expansion & evaluation, and system presentation. At the data acquisition & preprocessing stage, all military phrases and words from the National Military Dictionary and articles of the Cutting-edge Military Database were collected and transferred into a database for further evaluation. The ontology building stage adopted the FCA method and process to automatically build up the ontology structure in the military field by integrating and correlating the National Military Dictionary with the Cutting-edge Military Database. In the query expansion & evaluation stage, the K2FCQE is proposed and cross-analyzed with two other existing methods: the FCA-based query expansion and K2CM query expansion. In the first to third stages, we focused on the design and assessment of the query expansion algorithm. In the fourth stage, we developed a prototype of the Military News Retrieval System to recommend the derivative query concept to users, which will be revealed in the next section.

The research framework is shown in Fig. 1, whose steps are discussed from the perspective of performing empirical tasks.

3.2 Research Processes

3.2.1 Data acquisition & preprocessing

There are two sources of data collection for this research: the National Military Dictionary that is edited for various units of military professionals, with military terminology edited serving as the foundation of ontology building, and the database of news articles from the Defense Technology Monthly (DTM) magazine, a professional magazine in the military field that has a website dedicated to important military history that can be combined with military

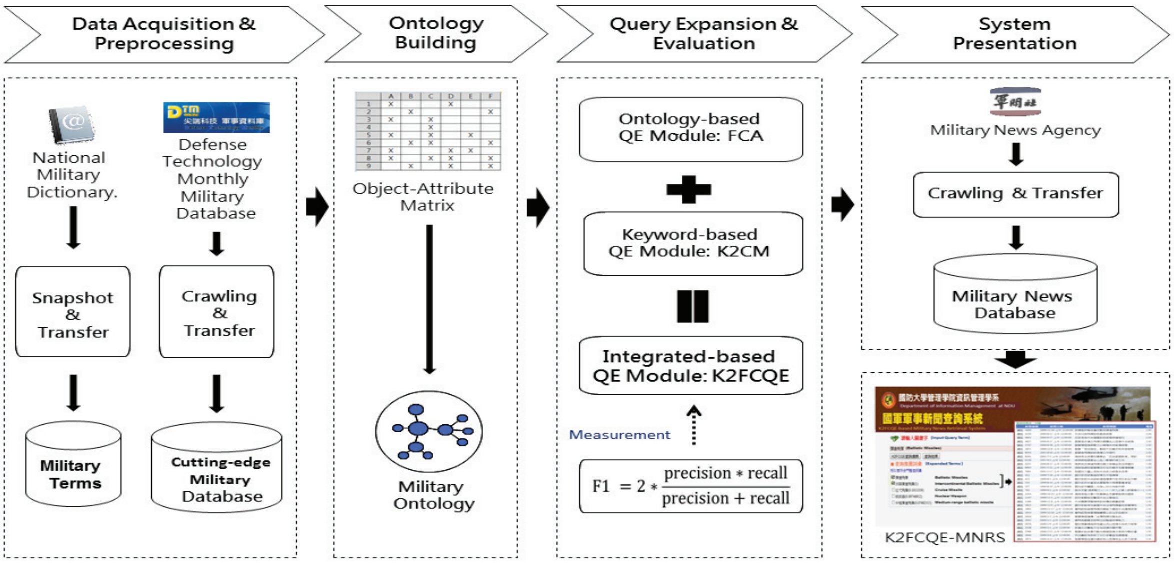

Fig. 1. Research framework

terminology to build a military knowledge database.

- **Military terminology**

Military terminology is a specialized set vocabulary that expresses military concepts encompassing the entire range of military thinking and philosophy. An investigation into the National Military Dictionary issued by the Ministry of National Defense found that the military terminology can be divided into 10 chapters—including national defense; war and strategy; political battle; personnel and military service; information; battles; training; rear service; communication, electronic warfare and electronic resources; and R&D on defense technology and weapons—with a total of 6,229 military terms. This research aims to develop a program to automatically retrieve military terminology and transfer it to the system to establish a military ontology.

- **Military reports**

Military reports are professional military documents that provide important sources for users to understand the developments in the military field. DTM (http://www.dtmonline.com), founded in January of 1984, is a well-respected professional military journal that reports on the weaponry of all countries, the latest technological developments of national defense, and domestic and international military news. It collects all the important military news and establishes a website for users to search for full-text articles. The journal database contains 24 military categories, including military equipment, the People's Liberation Army, aerospace science, ships, vehicles, artillery, and missiles. A total of 5,048 military news articles from 1990 to 2013 were collected to build up the ontology for the research.

3.2.2 Ontology building

In this study, the correlation between the attribute concept and object concept was established by an FCA process, and a prototype of military ontology was further built. The procedures and calculations were as follows:

(1) First, define the relationship between the attributes (military terminology) and objects (military report).
(2) Form a binary relation matrix of file and word and cross-examine it with the terminology and report one by one through an automatic program; if a file contains the target word, it is marked as "X," and then it will form a binary matrix with file and word. The military report and collection are defined as G for object, and military terminology as M for attributes, from which the context C can be induced, and the correlation between file and word is I; the formula is C = (G, M, I).

(3) Define A as part of G, while B as part of M; therefore, A ⊆ G; B ⊆ M. All the conditions that meet A = B' and B = A' are concept C, presented as C(A,B). A' represents all the common words of A while B' represents all the common files of B. In other words, among the binary matrix of file and word, the most common word in the largest file is a concept, and all the concepts in the standard context will be discovered through the prescribed formula.
(4) Calculate the hierarchy between concepts: D is defined as the file collection in the study and T as the word collection, such that if the relation of $(D_1,\ldots,D_m) \subseteq (D_1,\ldots,D_m,\ldots,D_n)$ exists in concept $\{D_1,D_2,\ldots,D_m,\ldots,D_n,T_i\}$ and concept $\{D_1,D_2,\ldots,D_m,T_j\}$, then the word collection T_i is the father of T_j; however, if the word collection T_j in all file collections is not fully covered by T_i, there is an interlaced relation between T_j and T_i; when word collection T_j in all file collections T_i is irrelevant, then T_j and T_i are independent. Repeat the above rule until all the relations between concepts are determined.
(5) Use the FCA Concept Explorer to transfer the context of the military files and terminology into the grid interface.
(6) Establish the hierarchy and interrelations between concepts so that the concept grid can be produced to represent military knowledge; the father node and child node in the grid concept can be reserved in the military ontology for further query expansion.

The chapter of the National Defense Technology and Weapon can serve as an example. There are 597 technical military terms that are treated as attributes, and 5,342 military news reports are represented as objects. All the technical terms are cross-analyzed one-by-one with the news articles; if a news article contains the query word, it will be marked as "X," and then a binary matrix will be formed between the file and word, which is the formal context of military file and military terminology.

3.2.3 Query expansion module: keyword to formal concept query expansion

Here, we introduce the modified query expansion method, K2FCQE that integrates the ontology and word concept to recommend query expansion.

Suppose D is the set of documents with $d_j(j=1\ldots, m)$; C is the set of concepts with $c_i(i=1\ldots, n)$; Q is the set of query terms with $q_k(k=1\ldots, k)$; and EC_i is regarded as expansion concepts with the set of recommendation terms $\{c_1, c_2, \ldots, c_i\}$. The mechanism and process are defined as follows:

(1) Empty all the query expansion words meant to be recommended to the user. EC is defined as a set of expansion concepts.

$$EC_i = \emptyset;$$

(2) Calculate the query word q_k and each concept c_i in the conception collection C of the ontology.

(2.1) Calculate the attaching weight $aw_{k,i}$ of q_k, which is the association degree between q_k and c_i.

$$aw_{k,i} = \log\left(\frac{N}{n_k}+1.0\right)\cdot tf_{k,i}\cdot \log\left(\frac{l_{k,i}}{l_i}+1.0\right) \quad (2)$$

where n_k is number of concepts associated with q_k; l_i is number of documents labeled by concept c_i; $l_{k,i}$ is number of documents including the term q_k and simultaneously labeled by concept c_i; $tf_{k,i}$ is regarded as the term frequency.

$$tf_{k,i} = \sum_{d_j \in D_i} \frac{\mathrm{count}(q_k; d_j)}{len(d_j)} \quad (3)$$

where D_i is a set of documents labeled by concept c_i; d_j is an instance in concept c_i; count $(q_k;\ d_j)$ is the frequency of q_k appearing in document d_j; len(d_j) is the document length of d_j.

(2.2) Calculate the $cw_{k,i}$ of q_k, which is the co-occurrence value of q_k and c_i within an effective word context.

$$cw_{k,i} = \frac{tpf_{k,i} \cdot \log\left(\frac{m_{k,i}}{M} + 1.0\right)}{\log(avgdist_{k,i} + 1.0)} \tag{4}$$

where $m_{k,i}$ is number of term-concept pair (q_k, c_i) ; $avgdist_{k,i}$ is the average distance between window size W and term-concept pair (q_k, c_i); $tpf_{k,i}$ is regarded as term-concept pair frequency.

$$tpf_{k,i} = \frac{\text{count}(q_k; c_i, W)}{Max_{j=1\ldots N}(\text{count}(q_k; c_j, W))} \tag{5}$$

where $\text{count}(q_k; c_i, W)$ is the number of term-concept pair (q_k, c_i) scanning with window size (W) in the document set D.

(2.3) Multiply $aw_{k,i}$ and $cw_{k,i}$ to get keyword-concept association $kca_{k,i}$ and save the result in collection ak_j.

$$kca_{k,i} = cw_{k,i} \times aw_{k,i} \tag{6}$$

$ak_j = \{ak_1, ak_2, ak_3, \ldots ak_{i-1}, ak_{i+1}, \ldots, ak_n\}$;
where $j \neq i$

(3) Normalize the $tca_{k,i}$ of ak_j by min-max normalization so that the correlation value is between 0 and 1.

$$ak_j = \frac{ak_j - \min(ak_j)}{\max(ak_j) - \min(ak_j)} \tag{7}$$

(4) Select the most highly related concept c' (i.e., max($kca_{k,i}$)) as the initial concept.
(5) Further calculate after removing the non- c' concept (c_i) from the collection.
(6) Calculate the similarity value s(c', c_i) of c' and non-c' concept with Jaccard index, and save the result in collection s_j . Moreover, as w = 0.5, objects (A) and attributes (B) are equally important in our experiments.

$$\text{S}(C_1, C_2) = w * \frac{|A_1 \cap A_2|}{|A_1 \cup A_2|} + (1-w)\frac{|B_1 \cap B_2|}{|B_1 \cup B_2|} \tag{8}$$

$s_j = \{s_1, s_2, s_3, \ldots s_{i-1}, s_{i+1}, \ldots, s_n\}$;
where $j \neq i$

(7) Multiply ak_j and s_j and save the result in collection QE_i. Rank the similarity of QE_i from high to low.

$$QE_i = ak_j \times s_j \tag{9}$$

(8) Capture the concept QE_i, which has higher similarity than the average threshold (α), and save it in the collection EC_i, and then recommend the result to the user. The computation of average threshold is shown in the section 4.1.2.

IV. RESEARCH FINDINGS

Based on the previous section, we construct a K2FCQE-MNRS and analyze our research findings.

4.1 Result Evaluation and Analysis

4.1.1 Module evaluation

To measure the effectiveness of K2FCQE, Boonchom and Soonthrnphisaj [26] and Bobadilla et al. [33] suggested three indicators (precision, recall, F-measure) that emphasize the quality of the set of recommended results [34]. In this study, we considered these indicators as evaluation measures of the query expansion performance. Precision indicates how many of the returned military reports are correct and helps the user filter irrelevant results. The recall value specifies how many of the military reports should have been found, and it helps the user discover all the relevant results. The F-measure is a summarized metric that combines precision and recall into a single value. The average performance is better when the F-measures are greater.

The evaluation metrics depend on the correct or relevant reports or documents that answer the expanded query words. Based on the preceding definitions and discussion, the formulae are stated below.

$$Precision = \frac{Number\ of\ correct\ military\ reports\ returned}{Number\ of\ military\ reports\ \ returned} \quad (10)$$

$$Recall = \frac{Number\ of\ correct\ military\ reports\ \ return}{Number\ of\ military\ reports\ in\ the\ specified\ ta} \quad (11)$$

$$F - measure\ = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (12)$$

The evaluation of retrieval performance is processed as follows:

Step 1: Retrieve the pre-categorized military reports from the Cutting-edge Military Database, including missiles, warships, and aerospace as the evaluation baseline.

Step 2: Use the terminology in the chapter of the National D efense Technology and Weapon as the query word.

Step 3: Conduct query expansion using a different model after keying in the query word and add all the recommended concepts one by one into the original query, and then calculate the recall, precision, and F-measure of the result. In the study, it is assumed that the F-measure will increase continuously; if it decreases, it means the retrieval performance degrades and the query expansion ceases; thus, the recall, precision, F-measure, and the numbers of expanded words as well as the similarity value of the last concept expansion word will be recorded.

Step 4: Add all the recall, precision, and F-measures of the query when the expansion ceases, and divide the sum by the total query words to get the average retrieval performance. The pros and cons of different query expansion models can then be analyzed.

Take Table 1 for example; when the user keys in “anti-submarine aircraft,” the F-measure is 0.1308; after adding the first concept in word “submarine” to the query, the F-measure increases to 0.2588; at this time, the query expansion continues as the value increases. Until adding the fourth concept word, which is “frigate,” the F-measure decreases to 0.2730, lowering the third concept word “torpedo” valued at 0.2776; at this time, the query expansion stops and the F-measure of the third query expanded concept will be recorded.

Table 1. Retrieval result of the query word “anti-submarine aircraft”

Query Word - Expanded Concept Word	Recall	Precision	F-measure
Antisubmarine Aircraft	0.1434	0.1203	0.1308
Anti-submarine Aircraft - Submarine	0.5902	0.1657	0.2588
Anti-submarine Aircraft - Submarine, Destroyer	0.7951	0.1664	0.2752
Anti-submarine Aircraft - Submarine, Destroyer, Torpedo	0.8443	0.1661	0.2776
Anti-submarine Aircraft - Submarine, Destroyer, Torpedo, Frigate	0.8934	0.1611	0.2730

4.1.2 Results analysis and discussion

The baseline was established before the evaluation. Therefore, the military news report on missiles, warships, and military aerospace from the DTM website were adopted as the performance evaluation baseline. There were 537 military aerospace reports, 244 warship reports, and 112 missile reports.

The military terminology in the chapter of National Defense Technology and Weapons was used to simulate the user’s query words (Table 2), among which 26 aviation and space related words corresponded to the military aerospace category of DTM, 108 rockets and missile-related words corresponded to the missile category, and 55 sea-weapon-related words corresponded to the warship category; the total number of experimental query words were 189.

The research analyzed the 189 military technical words to determine the retrieval performance of the four query expansion computing models discussed in the previous sections: No query expansion (keyword-based, without query expansion), FCA-based [28], K2CM [32], and K2FCQE (this study). Because of the limited length of the report, the results are only partially listed in Table 3.

Table 2. Simulated query word (partial)

Chapter	Classification	Terms No.	Terminology (examples)
Aviation & Space	Aerospace	26	Warning aircraft, electronic warfare aircraft, communications relay aircraft, air refueling, trainer, etc.
Rocket & Missile	Missile	108	Rockets, single rockets, multi-stage rockets, missiles, strategic missiles, tactical missiles, cruise missiles, etc.
Sea Weapon	Warship	55	Aircraft carriers, cruisers, battleships, frigates, destroyers, submarines, etc.

Fig. 2 presents the average comparison of the four models by precision, recall, and F-measure. The data show the following:

- K2FCQE was not only evaluated with three important indicators but also computed the similarity between the query word and recommended terms. This study adopts the average similarity as the threshold to determine the expanding query words. Through our experiments, we established that the threshold is equal to 0.065855. Based on this threshold, we further constructed a comprehensive module and a prototype system to retrieve military news.
- In general, the query expansion aims to increase the retrieval results of files for the FCA-based, K2CM, or K2FCQE model; it also effectively increased the recall rate of files, and although the precision was relatively low, the average retrieval performance was still more than twice that of the No query expansion model, judging from the F-measure scores.
- The average number of query expanded words processed by the K2FCQE was the highest among the three query expansion models, and the average F-measure and recall was the highest, which demonstrates that the K2FCQE model proposed in this study is superior to the other three models.

Compared to other related research on retrieval performance, the evaluation figure was relatively low in our study. There are three possible reasons for this phenomenon: (1) the terminologies of the National Military Dictionary were reviewed and revised by experts after discussion; on the contrary, the military report from the Cutting-edge Military Database were diversified and rapidly changing. As a result, some military terms may become obsolete or have no corresponding concepts in military reports; (2) the DTM database is classified in an artificial manner, which could lead to classification deviations; (3) there is no academically certified standard for evaluating the retrieval performance in the domain of defense and military knowledge.

However, the retrieval evaluation can still be applied as a reference to analyze the advantages and disadvantages of different retrieval models.

4.2 System Presentation

4.2.1 Military news articles

The MNA is affiliated with the Department of Defense and is responsible for military news interviews and reports. All the military news within this site are experimental sources of this study. The main task of the MNA is to provide military-related news, pictures, and videos to domestic and international media. The MNA provides not only a variety of military news but also uses many channels (e.g., Internet, satellite communications) to swiftly and rapidly publish military news. This study designed a web crawler module to extract 44,176 military news items between 2000 and 2011 from the MNA website. The format transfer module could remove HTML tags of military news pages and import the news content into the database.

4.2.2 K2FCQE-based military news retrieval system

For enabling users to acquire more query-related military news, this study intends to construct a K2FCQE-MNRS using ASP.NET and C#. This system can recommend more derivative concepts to the user by employing the similarity between the query term and ontology relationships, which in turn can help users to strengthen their comprehension of knowledge from the military domain.

Table 3. The retrieval performance of the four models (partial)

Query Word	No QE				Alqadah and Bhatnagar [28]: FCA				Tian et al. [32]: K2CM				Our proposed method: K2FCQE				
	Num/QE	Recall	Precision	F-measure	Num/QE	Recall	Precision	F-measure	Num/QE	Recall	Precision	F-measure	Num/QE	Recall	Precision	F-measure	Similarity
1. Anti-submarine Aircraft	0	0.14344	0.12027	0.13084	2	0.72950	0.17131	0.27747	3	0.66393	0.17344	0.27504	3	0.84426	0.16612	0.27762	0.178895
2. Warning Craft	0	0.13966	0.20107	0.16483	2	0.36312	0.23297	0.28384	5	0.48417	0.15767	0.23787	6	0.76350	0.13698	0.23229	0.108381
3. Medium-range ballistic missile	0	0.05357	0.12000	0.07407	4	0.26785	0.11538	0.16129	6	0.33035	0.11280	0.16818	4	0.44642	0.09823	0.16103	0.028255
…….	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
188. Surface to Surface Missile	0	0.09821	0.13253	0.11282	3	0.34821	0.10627	0.16283	4	0.45535	0.09659	0.15937	4	0.46428	0.09285	0.15476	0.001314
189. Amphibious Transport Dock Ship	0	0.02868	0.50000	0.05426	1	0.07377	0.31034	0.11920	5	0.42213	0.18968	0.26175	6	0.85245	0.16547	0.27714	0.159735
Average	0	0.10110	0.24145	0.07691	3.1707	0.24602	0.20632	0.14295	2.9758	0.27131	0.18473	0.14723	3.3577	0.38612	0.17572	0.16720	0.0658551

*Num/QE: The expanded query word numbers when the query expansion stops.

	Recall	Precision	F-measure
No QE	0.1011	0.24145	0.07691
Alqadah & Bhatnagar [28]: FCA	0.24602	0.20632	0.14295
Tian et al. [32]: K2CM	0.27131	0.18473	0.14723
Our Proposed Method: K2FCQE	0.38612	0.17572	0.1672

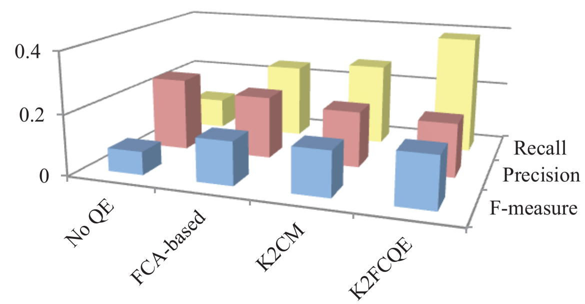

Fig. 2. The retrieval performances of the four models

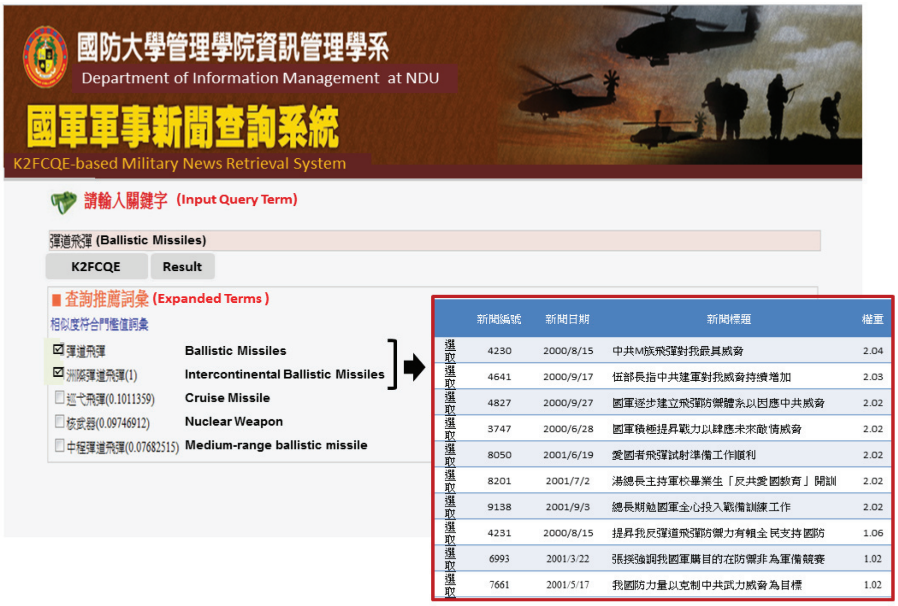

Fig. 3. The user interface and query results of K2FCQE-MNRS (example for "Ballistic Missiles")

This study adopted the MNA website as the data source and K2FCQE as the query expansion approach to develop a prototype for a Military News Retrieval System. Taking the query term "Ballistic Missiles" as an example, K2FCQE-MNRS used similarity ranking to recommend four expanding terms to users: International Ballistic Missiles, Cruise Missile, Nuclear Weapon, and Medium-range Ballistic Missiles. Users can select expanded terms recommended by K2FCQE-MNRS to retrieve the relevant military news content (see Fig. 3).

This system can recommend more related concepts to users that can help them to search for derivative military news more effectively.

V. CONCLUSION

In this study, the military vocabulary from the National Military Dictionary and military reports from the Cutting-edge Military Database were used as data sources to build a military ontology through methods of formal conceptual analysis. Moreover, a hybrid query expansion, K2FCQE, was introduced as an integration of the FCA-based retrieval with the K2CM. The experimental results indicate that the performance of K2FCQE was superior to the other query expansion methods.

Furthermore, this study developed a retrieval system of military news that provides more query-related terms to users. There are two main contributions of this research. First, the automatic building of a military ontology using the FCA method to improve the efficiency of the traditional ontology was built in a manual or semi-automatic way. Second, the proposal of a hybrid concept, K2FCQE, and its tested effectiveness in a comparison analysis with other methods. Besides, a Military News Retrieval System based on K2FCQE method, K2FCQE-MNRS, was constructed to expand users' query terms in a military-related domain.

The remainder of this section will present recommendations for future research. First, the vocabulary or dictionary used in this study was limited by the scope of the study as there are inherent restrictions in terms of time and space with a dictionary considering that the ontology will remain unchanged and there may be a loss of information if new vocabulary is added. This may give rise to the need of directing new vocabulary items back into the ontology to sustain itself to stay updated in the field. Next, this study was based on expert analysis for processing professional vocabulary, but common users may not be as aware of the concepts as military experts and scholars are in this field. More user-friendly information can be provided if individual concepts or a collaborative recommendation module is added to the proposed model to perform the analysis of query expansion.

ACKNOWLEDGEMENT

This research was partially sponsored by the Ministry of Science and Technology (MOST), Taiwan under Grant No: 102-2410-H-606 -007 & 103-2410-H-606 -005 -MY2

REFERENCES

[1] Liu, X., and Fang, H., "Latent Entity Space: a Novel Retrieval Approach for Entity-Bearing Queries," Information Retrieval Journal, Vol. 18, No. 6, pp. 473-503, 2015.

[2] Jovanavić, J., Devedžić, V., Gaševič, D., Brooks, C., Hatala, M., Eap, T., and Richards, G., "Using Semantic Web Technologies to Analyze Learning Content,"IEEE Internet Computing, Vol. 11, No. 5, pp. 45-53, 2007.

[3] Swartout, B., Patil, R., Knight, K., and Russ, T., "Toward Distributed Use of Large-Scale Ontologies," In Proceedings of the 10th Workshop on Knowledge Acquisition for Knowledge-Based Systems, Banff, Canada, November, pp. 138-148, 1996.

[4] Conesa, J., Storey, V. C., and Sugumaran, V., "Improving Web-Query Processing through Semantic Knowledge," Data & Knowledge Engineering, Vol. 66, No. 1, pp. 18-34, 2008.

[5] Hachey, B. Grover, C. Karkaletsis, V. Valarakos, A. Pazienza, M.T. Vindigni, M. and Coch, J. "Use of Ontologies for Cross-Lingual Information Management in the Web," In Proceedings of the Ontologies and Information Extraction International Workshop Held as Part of the EUROLAN, Bucarest, Romania, July 28- August 8, 2003.

[6] Chen, L. C., Chao, W. T., and Hsieh, C. J., "A novel query expansion method for military news retrieval service," In International Conference on Asian Language Processing (IALP), Kuching, Malaysia, October 20-22, 2014.

[7] Spink, A., Wolfram, D., Jansen, M. B., and Saracevic, T., "Searching the Web: The Public and Their Queries," Journal of the American society for Information Science and Technology, Vol. 52, No. 3, pp. 226-234, 2001.

[8] Rocchio, J., "Relevance Feedback in Information Retrieval," In the Smart Retrieval System: Experiments in Automatic Document Processing, G. Salton, Prentice-Hall, Englewood Cliffs, NJ, pp. 313-323, 1971.

[9] Wu, I. C., Chen, G. W., Hsu, J. L., and Lin, C. Y., "An Entropy-Based Query Expansion Approach for Learning Researchers' Dynamic Information Needs," Knowledge-Based Systems, Vol. 52, pp. 133-146, 2013.

[10] Zheng, H. T., Borchert, C., and Kim, H. G., "Exploiting Gene Ontology to Conceptualize Biomedical Document Collections," In Proceedings of the 7th International Semantic Web, Karlsruhe, Germany, Springer, October 26-30, pp. 375-389, 2008.

[11] Efthimiadis, E. N., "Query Expansion," Annual Review of Information Science and Technology, Vol. 31, pp.121-187, 1996.

[12] Qiu, Y., and Frei, H. P., "Concept Based Query Expansion," In Proceedings of the 16th annual international ACM SIGIR Conference on Research and Development in Information Retrieval, Pittsburgh, PA, USA, ACM, June 27-July 1, pp. 160-169, 1993.

[13] Segura, N. A., Garcia-Barriocanal, E., and Prieto, M., "An Empirical Analysis of Ontology-Based Query Expansion for Learning Resource Searches Using MERLOT and the Gene Ontology," Knowledge-Based Systems, Vol. 24, No. 1, pp. 119-133, 2001.

[14] Jiang, G., Ogasawara, K., Endoh, A., and Sakurai, T., "Context-Based Ontology Building Support in Clinical Domains Using Formal Concept Analysis," International Journal of Medical Informatics, Vol. 71, No. 1, pp. 71-81, 2003.

[15] Wille, R., "Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts," In Ordered Sets, Netherlands, Springer, pp. 445-470, 1982.

[16] Formica, A., "Ontology-Based Concept Similarity in Formal Concept Analysis," Information Sciences, Vol. 176, No. 18, pp. 2624-2641, 2006.

[17] Uta, P., “Formal Concept Analysis in Information Science,” Annual Review of Information Science and Technology Vol. 40, pp. 521-543, 2006.
[18] Ganter, B., and Wille, R., Formal Concept Analysis: Mathematical Foundations, Springer Science & Business Media, 2012.
[19] Stumme, G., “Ontology Merging with Formal Concept Analysis,” In the Proceedings of the Dagstuhl Seminar 04391 on Semantic Interoperability and Integration, Germany, September 19-24 2005.
[20] Gruber, T. R., “A Translation Approach to Portable Ontology Specifications,” Knowledge Acquisition, Vol. 5, No. 2, pp.199-220, 1993.
[21] Van Heijst, G., Schreiber, A. T., and Wielinga, B. J., “Using Explicit Ontologies in KBS Development,” International Journal of Human-Computer Studies, Vol. 46, No. 2, pp. 183-292, 1997.
[22] Corcho, O., Fernandez-Lopez, M., and Gomez-Perez, A., “Methodologies Tools and Languages for Building Ontologies. Where is Their Meeting Point,” Data & Knowledge Engineering, Vol, 46. No. 1, pp. 41-64, 2003.
[23] Noy, N. F., and Mcguinness, D. L., “Ontology Development 101: A Guide to Creating Your First Ontology,” Stanford Medical Informatics Technical Report SMI-2001-0880, 2001.
[24] Bhogal, J., Macfarlane, A., and Smith, P., “A Review of Ontology Based Query Expansion,” Information Processing & management, Vol. 43, No. 4, pp. 866-886, 2007.
[25] Voorhees, E. M., “Query Expansion Using Lexical-Semantic Relations,” In SIGIR’94, London, Springer, pp. 61-69, January 1994.
[26] Boonchom, V. S., and Soonthornphisaj, N. “ATOB Algorithm: an Automatic Ontology Construction for Thai Legal Sentences Retrieval,” Journal of Information Science, Vol. 38, No. 1, pp. 37-51, 2012.
[27] Yokoyama, A., and Klyuev, V., “Search Engine Query Expansion Using Japanese WordNet,” HumanCom 2010, Cebu, Philippines, August 11-13, 2010.
[28] Alqadah, F., and Bhatnagar, R., “Similarity measures in formal concept analysis,” Annals of Mathematics and Artificial Intelligence, Vol. 61, No. 3, pp. 245-256, 2011.
[29] Lee, M. C., Chen, H. H., and Li, Y. S., “FCA based Concept Constructing and Similarity Measurement Algorithms,” International Journal of Advancements in Computing Technology, Vol. 3, pp. 97-105, 2011.
[30] Weng, S. S., Tsai, H. J., Liu, S. C., and Hsu, C. H., “Ontology construction for information classification,” Expert Systems with Applications, Vol. 31, pp. 1-12. , 2006
[31] Zhao, Y., and Halang, W., “Rough Concept Lattice Based Ontology Similarity Measure,” In Proceedings of the 1st International Conference on Scalable Information Systems, Hong Kong, ACM, May 29-June 1, 2006.
[32] Tian, X., Du, X. Y., and Li, H. H., “Computing Term-Concept Association in Semantic-Based Query Expansion,” Journal of Software, Vol. 19, No. 8, pp. 2043-2053, 2008.
[33] Lu, S., and Bai, S., “Quantitative Analysis of Context Field in Natural Language Processing,” Chinese Journal of Computers, Vol. 24, No. 7, pp.742-747, 2001.
[34] Bobadilla, J., Ortega, F., Hernando, A., and Gutiérrez, A., “Recommender Systems Survey,” Knowledge-Based Systems, Vol. 46, pp.109-132, 2013.