A Web mining-based Case Adaptation Model for Quality Assurance of Pharmaceutical Warehouses

Abstract

Quality assurance (QA) of pharmaceutical warehouses is a critical issue for formulating working instructions in accord with many regulations. Currently, such QA planning, relying on human experience, not only increases the chance of errors and quality problems for pharmaceutical products, but also results in high rejection and recall rates. This paper proposes a web mining-based quality assurance system (WMQAS), integrating case-based reasoning (CBR) and a hybrid web mining technique, to facilitate decision making in QA operations in pharmaceutical warehouses. CBR is applied for solving complex new problems based on past cases, however, current case adaptation processes are mainly based on the experience of expert where past solutions are modified. The proposed system embeds a hybrid web mining technique in the case adaptation engine to extract the useful web information for decision makers. Consequently, the effectiveness and efficiency of QA processes can be enhanced, and the product quality can be maintained.

Keywords - Case-based reasoning, web mining, quality assurance, pharmaceutical products

1. Introduction

Quality assurance (QA) plays an important role in all industries, especially in handling pharmaceutical products (Alli, 2016; Jie et al., 2016; Shibata et al., 2016; Newton, 2015; Aung & Chang, 2014). Characterized by high product variety, smaller product size and environmental sensitivity, ensuring the quality of pharmaceutical products places significant pressure on logistics service providers (LSPs) in handling such products in warehouses. LSPs are currently striving to develop appropriate and effective QA schemes to reduce product discrepancies as well as rejection and recall rates. Conventionally, LSPs consolidate and store various types of pharmaceutical products in warehouses and outsource secondary packaging processes to third-party packaging companies (Olah et al., 2017; Mangan & Lalwani, 2016; Rossetti et al., 2011). After the packaging operations, the products are sent back to the warehouses before delivering to customers. This is because carrying out secondary packaging processes requires compliance with Good Manufacturing Practice (GMP) without which current LSPs are not allowed to perform such processes in their warehouses. This outsourcing process not only requires additional time and cost in moving and handling pharmaceutical products, but also increases the chance of errors due to poor communication between the parties involved. Therefore, LSPs tend to carry out such packaging processes in their own warehouses rather than outsourcing to

third party packaging companies (Quinn, 2013). Differing to general goods, more attention needs to be paid in each step of the packaging process, storage environment, equipment, staffing and facilities so as to fulfill the requirement of GMP and ensure the quality of such products. For example, confusion can easily occur in labeling materials due to the high similarities of product names, dosages and ingredients, which can be disastrous or even fatal to the end user if anything goes wrong. Any failure in the packaging of such products could result in a quality change which may result in either a failure to cure, injury, illness or even fatal incidents to the patients (Papalexi et al., 2016; Paine & Lockhart, 2012).

Currently, there are two major problems in implementing QA in pharmaceutical warehouses, as shown in Figure 1. Firstly, LSPs are required to manually formulate an appropriate QA scheme which generally contains many policies, working instructions and procedures, with detailed explanations for preventing incidents occurring in the warehousing and packaging processes (Heinrich, 2015; Lawrence et al., 2014). However, such processes are complicated due to the need to review a number of international standards and customers' specifications, and the likelihood of human errors in handling numerous documents. Implementing an inadequate QA scheme may lead to undesirable incidents, such as mislabeling and contamination of products and hence affect the product quality (Kumar & Jha, 2015). In considering the complexity in the QA scheme formulation process, a decision support system is needed to make use of professional knowledge and past experience for improving its effectiveness and product quality. Secondly, some LSPs formulate QA schemes based on their subjective judgement and experience rather than in considering the updated and relevant information from official websites. Even if they are required to find information such as recall news and safety alerts from the Internet, they need to spend a long period of time on non-core businesses, i.e. manually searching and analyzing the association of the information. At present, there is no systematic mechanism for LSPs to capture useful medical and quality information from international institutions and official electronic documents. In considering the amount and variety of knowledge on the Internet, inefficient knowledge acquisition processes may affect the availability and relevance of decision making in adapting useful knowledge. Therefore, the QA scheme cannot be effectively formulated and executed, resulting in product rejections and recalls.

To take web information into consideration during the QA formulation process, a web mining-based quality assurance system (WMQAS), integrating the CBR and a hybrid web mining technique, is proposed to formulate appropriate working instructions and procedures for QA schemes and to identify possible incidents in such operations. CBR allows the users to make use of their knowledge and past cases for solving new problems. In order to acquire online information for CBR, a hybrid web mining technique, i.e. term frequency-inverse

document frequency-based document clustering (TIDC) algorithm, is embedded in the case adaptation process. The effectiveness and efficiency of case adaptation in CBR can be thus enhanced. With the help of WMQAS, a proactive approach can be provided to identify potential incidents in warehousing and packaging operations so as to maintain product quality.

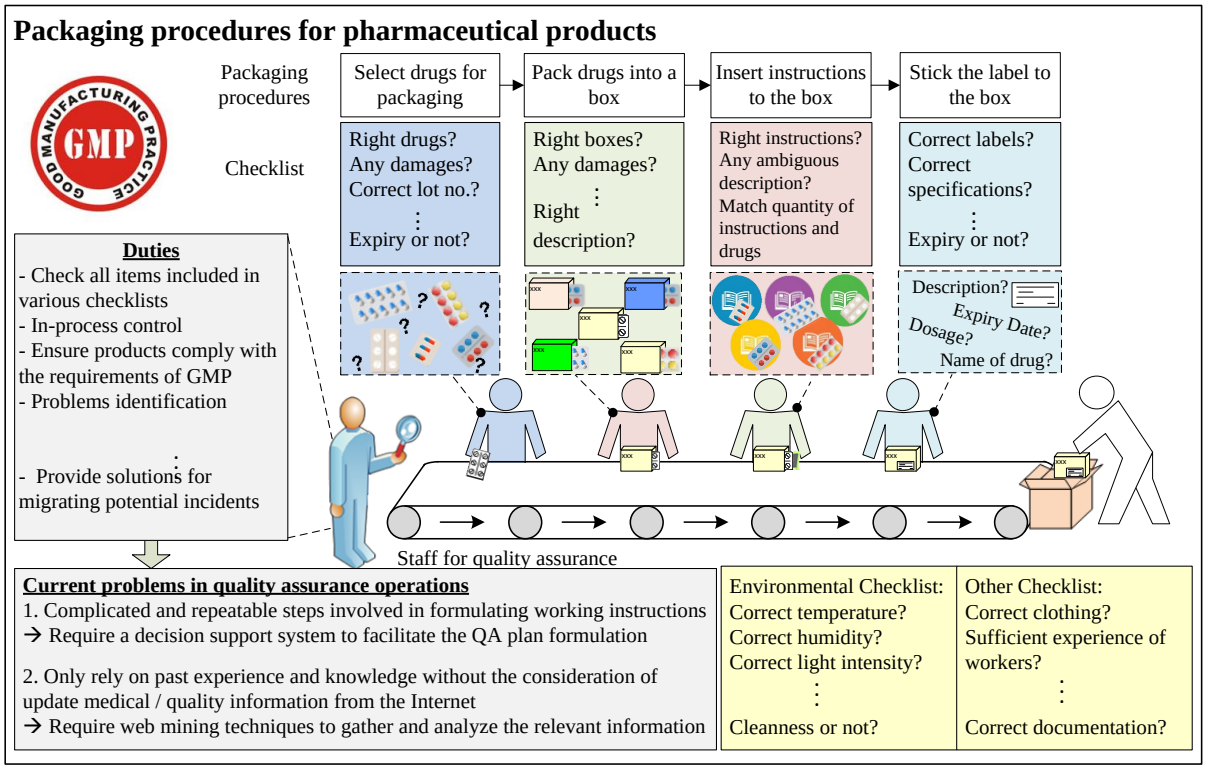


Figure 1. Packaging procedures for pharmaceutical products in warehouses

The rest of the paper is organized as follows. In Section 2, the literature related to the current situations of pharmaceutical warehouses, international standards, CBR and web mining techniques is reviewed. Section 3 describes the architecture of the proposed system while a case study is presented in Section 4 for illustrating the use of WMQAS in a real situation. Section 5 discusses the results and the advantages of launching the proposed system. Conclusions are presented in Section 6.

2. Literature Review

Considering the growing awareness of the significance of the quality of pharmaceutical products, effective QA schemes are of the upmost importance for LSPs (Sharifian et al., 2016; Woodcock, 2004). LSPs provide locations for consolidating and storing pharmaceutical products from various suppliers and in performing value adding services, such as palletization (Lim et al., 2013; Gu et al., 2007). In addition, they have to carry out additional secondary packaging services for such products. According to the Pharmacy and Poisons Board of Hong Kong (2014), secondary packaging is a manufacturing step for labelling, relabeling, cartoning, re-cartoning, or adding additional information to pharmaceutical products with the primary packaging materials. Implementing ineffective QA processes in warehousing and additional

packaging operations may cause the contamination of such products, non-conformity of labeling materials of pharmaceutical products and finally lead to product rejection and recall with serious impact on LSPs and patients.

In general, QA is a process to ensure that products are manufacturing with good quality and can satisfy customers (Robinson, 2017). The aim of QA is the conformance of production processes, products and services with international best practices, standards and requirements (Wingate, 2016; ASOSAI, 2009; Morenzo-Luzon & Peris, 1998). Currently, Good Manufacturing Practice (GMP), Good Distribution Practice (GDP) and the Food and Drug Administration (FDA) are common standards that are widely applied for controlling the quality of pharmaceutical products (Haleem et al., 2015). Through systematic measurement and standardization in process control, QA can not only detect variation of product and control quality, but also needs to emphasize the quality in each step of the production process (Velayutham, 2003). QA provides confidence to customers by continuous complying with the quality requirements (Mahzan & Hassan, 2015; Karapetrovic & Willborn, 2000). Therefore, researchers have focused on adopting different approaches in order to ensure the quality of such products (Marinkovic et al., 2016; Felice et al., 2014; Percin & Min, 2013; Anand et al., 2012; Gray et al., 2011). Yu (2008) proposed the concept of quality by design to ensure that pharmaceutical products had predefined quality through the understanding and control of manufacturing variables. Troup & Georgakis (2013) adopted process systems engineering tools for monitoring and control, modeling and optimization so as to reduce the risks to pharmaceutical products. EI-Menhawy et al. (2015) applied the six sigma methodology to control the variation of pharmaceutical tablets and to achieve the required quality during manufacturing. Although human based QA or decision making can help to manage operations in warehouse, misinterpretation and slackness in implementing QA may lead to inconsistent actions and result in poor operation performance (Seo et al., 2015). For instance, due to the environmental sensitivity characteristic of pharmaceutical products, LSPs should pay high attention to the storage environment so as to prevent the large fluctuation of temperature, humidity and light intensity, or even out of the required storage specifications. The slackness in that aspect may lead to product contamination, resulting in product rejection and recall. Moreover, the characteristics of pharmaceutical products, including high variety, small in size and first expired first out strategy, increase the difficulties for LSPs in implementing QA scheme manually and in identifying potential incidents according to international quality standards. To effectively implement QA and improve the quality of decision making, acquiring knowledge from past records becomes crucial in providing reliable references for LSPs to solve problems in particular situations.

Case-based reasoning (CBR) is one of popular artificial intelligence techniques that

emulates human thinking to facilitate decision making (Kolodner, 1993; Aamodt & Plaza, 1994). It makes use of knowledge by recalling and reusing similar previous experience to provide decision support in solving new problems (Riesbeck & Schank, 2013; Wang et al., 2013). With self-learning ability, CBR allows users to reduce the time for acquiring information and knowledge from prior situations, so as to avoid repeating the mistakes that occurred previously (He & Tian, 2017; Alptekin & Büyüközkan, 2011). Because of the advantages of CBR, it is widely adopted in various application areas such as new product development, quality assurance and control, improvement in operation efficiency and risk management (Yao et al., 2014; Li & Sun, 2011; Belecheanu et al., 2003). Lee et al. (2014) developed a knowledge-based ingredient formulation system in personal care industry so as to improve the communication and knowledge sharing among various stakeholders in new product development processes. Chakraborty & Boral (2017) adopted a CBR system to fulfill varying requirements of users and achieve maximum machining performance from machine tool selection. In the area of quality assurance and control, Lao et al. (2012) proposed a hybrid decision support system that integrated CBR and fuzzy logic to monitor the receiving operations and the storage conditions of food so as to maintain the food quality in warehouses. Lam et al. (2013) applied CBR to design a risk control and monitoring system to formulate a follow up plan for handling incidents in wine distribution hubs so as to reduce the risks of deteriorating quality. To further improve the performance of solutions generated from CBR, some research studies focused on investigation of the case adaptation process (Qi et al., 2017; Jalali & Leake, 2016; He et al., 2009; Haque et al., 2000). Typically, current methods for case adaptation in CBR can be classified into two categories: transformational and generative adaptations (Wilke & Bergmann, 1998). For transformational adaptation, a new solution for solving new problems is generated by adding, deleting and modifying the retrieved past case while generative adaptation makes use of domain knowledge for a problem solver to construct the solution from scratch. These two case adaptation methods require detailed knowledge from past cases and domain experts (Zhang et al., 2007). Past cases are treated as the main sources of knowledge for handling the new problems. Experts also play important roles in these knowledge-intensive adaptation methods for advancing the retrieved case and form a new solution based on their subjective judgement and personal experience. However, new knowledge adaptation is not accessible and available for experts in such situations. In addition, although the importance of expert knowledge has been identified in the case adaptation process, Chebel-Morello et al. (2013) presented that the current methods for knowledge elicitation and case adaptation in CBR have low efficiency. In their study, the time for an expert to carry out the steps in a CBR system was 37.7% and 45.9% of their total computational time, respectively. This implies that inefficient case adaptation in CBR may greatly increase the total time in formulating solutions to solve new problems, even if a past case is retrieved. Fuchs et al. (2014) also pointed out that, currently, there is a lack of a systematic approach in CBR for acquiring

new knowledge in case adaptation. In view of that, an effective mining approach for acquiring relevant and update knowledge in case adaptation of CBR is essential for supporting experts to modify retrieved cases. Instead of traditional approach in knowledge acquisition from consulting experts, methods of online knowledge acquisition are getting more attention in recent years for capturing new knowledge in revising the retrieved records (Reyes et al., 2015). In order to effective extract updated medical information from the internet, researchers considered adopting the web mining technique for delivering the highly correlated medical information to physicians (Ting et al., 2013).

Web mining is a process of extracting and discovering the patterns, trends, directions and rules from unstructured information on the Internet, such as text documents, HTML file, emails and messages (Kim et al., 2016; Min 2010). It is an automated technique to systematically and efficiently identify, integrate, manage and exploit knowledge from a vast number of web documents (Thomaz et al., 2017). It has been applied to assist decision making in various areas, such as topic and content discovery, customer relationship management and identifying user opinions (Law et al., 2017; Salloum et al., 2017; Abrahams et al., 2015). Typically, there are three categories of web mining: web content mining, web structure mining and web usage mining. Web structure mining emphasizes the hyperlink structure of the web to link the different objectives together (Kanathey et al., 2018). Web usage mining is used to predict users behavior when they are interacting with the WWW. Web content mining adopts the concept and principles of data mining to discover knowledge from text and media documents (Rao & Arora, 2017; Xu, et al., 2011). K-mean clustering is one of most popular web content mining techniques to group web documents into various clusters according to their similarity (Ghosh & Duhey, 2013) and is proven to be an effective way to produce good clustering results, suitable for clustering large sets of data and producing globular clusters (Na et al., 2010). Dunn (1973) introduced the fuzzy c-mean (FCM) clustering algorithm, which is an extension of k-mean clustering, to remove the limitation of k-mean clustering by only assigning the data into one cluster. It associates data with the degree of membership to the cluster that they belong to (Bezdek et al., 1984). Ozer (2005) applied the FCM clustering algorithm to Internet portals so as to determine associations among potential users and hence provide unique information to each cluster. Al-Ayyoub et al. (2015) adopted the FCM clustering algorithm to image the analysis area for diagnosing and detecting different diseases in patients. Since online healthcare information can be classified into more than one category, the FCM clustering algorithm was found to be an appropriate approach for knowledge retrieval. With the use of web mining techniques in the case adaptation of CBR, it is believed that the performance of searching and gathering up-to-date information can be increased in term of effectiveness and efficiency.

Firstly, LSPs are required to manually formulate an appropriate QA scheme which

generally contains many policies, working instructions and procedures, with detailed explanations for preventing incidents occurring in the warehousing and packaging processes (Heinrich, 2015; Lawrence et al., 2014). However, such processes are complicated due to the need to review a number of international standards and customers' specifications, and the likelihood of human errors in handling numerous documents. Implementing an inadequate QA scheme may lead to undesirable incidents, such as mislabeling and contamination of products and hence affect the product quality (Kumar & Jha, 2015).

From the above literature, it is clear that implementing an effective QA scheme is necessary to provide the working instructions during warehousing and additional packaging processes without which LSPs cannot ensure the quality of pharmaceutical products. However, due to the characteristics of pharmaceutical products, this increases the challenges for LSPs to manually formulate QA with detail working instructions and procedures. Thus, a web mining-based quality assurance system (WMQAS) is proposed in this paper to facilitate decision making in implementing the QA scheme in pharmaceutical warehouses. With the use of CBR, appropriate working instructions and procedures can be formulated effectively for preventing the incidents occurred. Moreover, although researchers highlighted the adoption of CBR can help to improve the decision making in formulating of QA scheme, traditional case adaptation process in CBR relying on expert knowledge is low efficiency. Therefore, an online knowledge acquisition approach, i.e. web mining technique, is embedded in the case adaptation process to effective extract the web information for revising the retrieved past QA records so as to complying with the pharmaceutical international standards in warehouses.

3. Methodology

In this section, a web mining-based quality assurance system (WMQAS) is described that provides appropriate working instructions and procedures, and identifies possible incidents when formulating an effective QA scheme so as to ensure the high quality of pharmaceutical products. Figure 2 shows the architecture of WMQAS which is divided into two tiers: (i) data collection module and (ii) web mining-based decision support module.

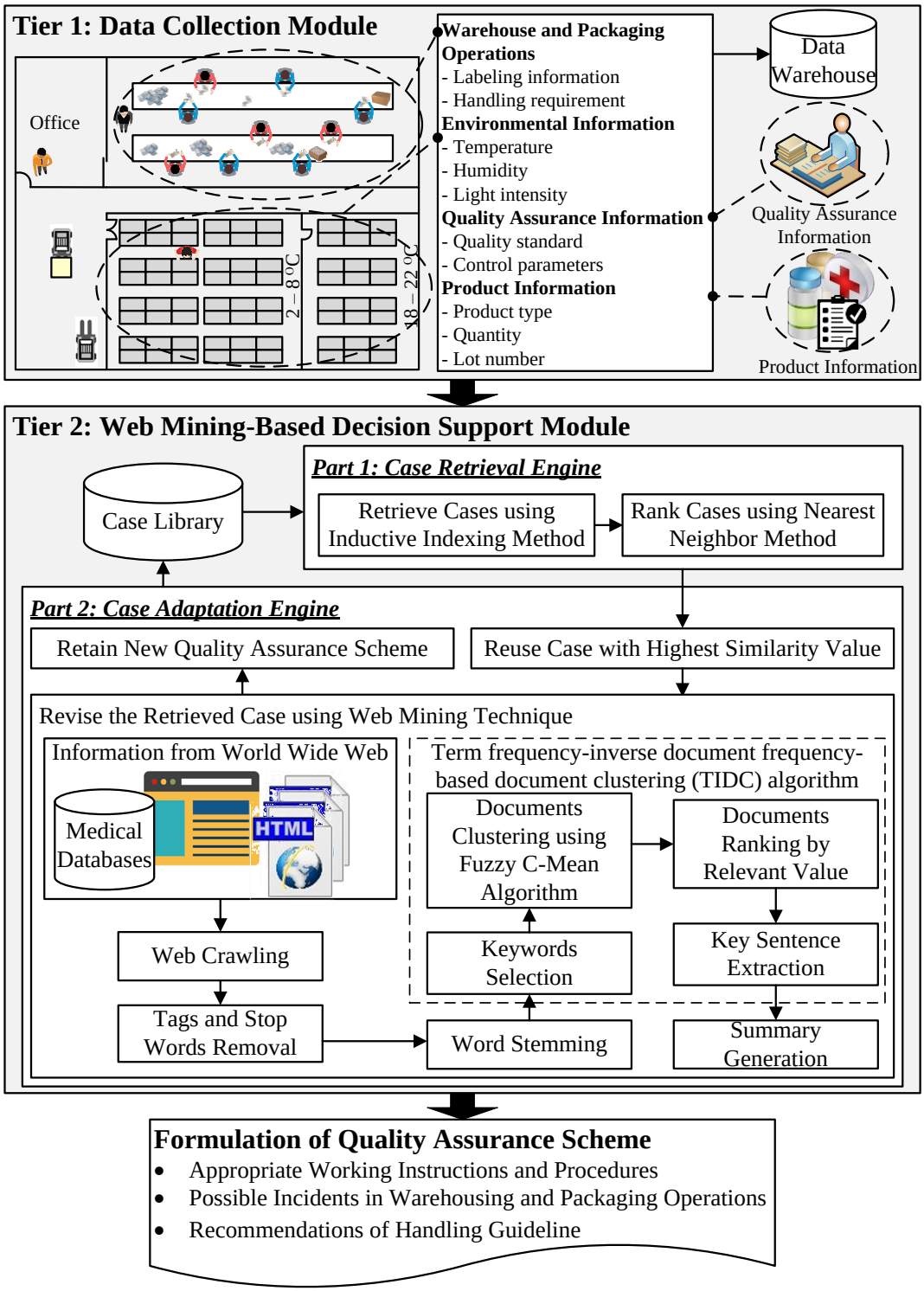


Figure 2. The architecture of the WMQAS

3.1 Tier 1 – Data Collection Module

3.1.1 Case Definition

Each case stored in the case library involves two parts which are the input problem and output solution parts, as shown in Figure 3. The case can be represented as a 7-tuple which is

{P, O, E, Q, I, A, R}. The first four tuples (P, O, E, Q) are input problems and the last three tuples (W, I, R) are the output solutions for the case where $P = \{P_1, P_2, \ldots, P_n\}$ is a set of basic information of products such as product name, type, expiry date, quantity for LSPs to distinguish particular services to be provided for customers; $O = \{O_1, O_2, \ldots, O_n\}$ stores a set of warehousing and packaging operations data such as ingredients and dosage in the product labelling, handling requirement, operation time, required staffing and equipment; $E = \{E_1, E_2, \ldots, E_n\}$ is environmental data that may affect the product quality including temperature, humidity, light intensity; $Q = \{Q_1, Q_2, \ldots, Q_n\}$ is a set of QA requirements for controlling the product quality at every step in the operations, such as quality guidelines, inspection elements and control parameters; $W = \{W_1, W_2, \ldots, W_n\}$ stores a set of working instructions and procedures for carrying the warehousing and packaging operations; $I = \{I_1, I_2, \ldots, I_n\}$ is a set of possible incident data that may occur in such operations; and $R = \{R_1, R_2, \ldots, R_n\}$ is a set of recommendations based on identified incidents for improving the product quality according to its quality standards. In order to construct the case library comprehensively, LSPs should identify a sufficient amount of successful QA records with recommendations for preventing the incidents occurring and store these QA records in the case library.

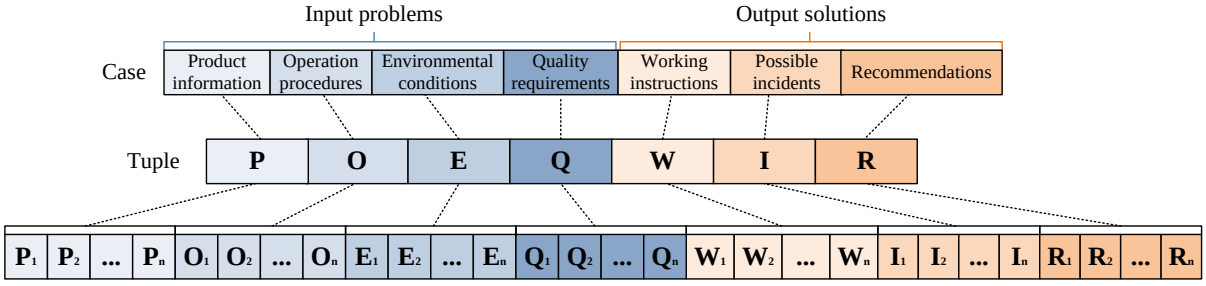


Figure 3. Structure of case stored in case library

3.1.2 Construction of Data Warehouse

Four types of input data in the problem part of the case, i.e. product information, warehousing and packaging operations, environmental conditions and QA requirement, are collected in this tier. Through the data pre-processing, incomplete and duplicated data are removed. Relevant data is then stored in the data warehouse for further processes in Tier 2.

3.2 Tier 2 – Web Mining-Based Decision Support Module

Once the LSPs execute QA processes in the warehouse, relevant data is passed to the web mining-based decision support module to support the decision making process in suggesting appropriate handling guidelines for mitigating potential incidents in such processes. Two case engines are involved in this module namely, (i) case retrieval engine and (ii) case adaptation engine.

3.2.1 Case Retrieval Engine

In order to retrieve the useful past QA records, LSPs have to identify the relevant key attributes according to the quality standards. These attributes, such as the difference between the required and actual temperature and regulation compliance, contain the data for examining the quality of pharmaceutical products at each step involved in warehousing and packaging operations. Based on the defined attributes, the case retrieval engine can search for relevant past QA records stored in the case library. Inductive indexing is applied to cluster the past QA records, where a decision tree is constructed based on defined attributes. By searching along each indexing level in a decision tree structure, a smaller set of potential QA records at the last level is selected.

With a list of potential QA records, the nearest neighbor method is employed to further analyze their suitability by calculating the similarity between the stored records and the new input case based on the weight features. The weight features show the importance and relationship to the product quality, ranging from 0 to 1. LSPs are required to define and input the corresponding features in the WMQAS so as to compute the total similarity value of the new case to past QA records based on Eq. (1).

$$similarity\ (case_{new}, case_{old}) = \frac{\sum_{i \in R} w_i \cdot sim(f_i^{new}, f_i^{old})}{\sum_i^n w_i} \quad (1)$$

where w_i is the weight of individual quality feature, *sim* is the similarity function of the feature and f_i^{new} and f_i^{old} are the values for feature i in the new case and the retrieved past QA records respectively. Records with different similarity values are than ranked in descending order. The most significant past QA record, the one with highest similarity value, is selected as a reference and transferred to the case adaption engine for generating the new solution.

3.2.2 Case Adaptation Engine

The retrieved past QA record with the highest similarity value is then reused in the case adaptation engine for solving the new problem. The traditional case adaptation process mainly relies on human experience to modify the retrieved case. Hence, updated medical and quality information for handling pharmaceutical products may be neglected, thus affecting the quality of the QA scheme. Thus, the web mining technique is adopted to improve the searching efficiency in capturing the relevant information from reliable sources in the Internet for revising the retrieved case.

A huge amount of information is available on the World Wide Web. With the process of web crawling, relevant web page and web content in particular databases can be gathered from time to time. Text pre-processing is employed after the web crawling to remove the stop words, such as a, the, of and in, and tags from documents in HyperText Markup Language (HTML), such as <html>, <head> and <body>. In addition, the word stemming step in text pre-

processing helps to reduce the derived and inflected words to their word stem, base or root form. For example, the words recall, recalls, recalling, recalled and recallable are mapped into the root word recall. Through text pre-processing, redundancy and noisy information can be removed from the web documents while the meaning of the web documents can be interpreted and transmitted effectively.

After pre-processing the web documents, the Term frequency-Inverse document frequency-based Document Clustering (TIDC) algorithm, integrated with the Term Frequency-Inverse Document Frequency (TF-IDF) weight method and FCM algorithm, is designed for assessing and weighting the importance of the keywords in individual documents and then cluster them into meaningful categories. The objective function of the TIDC algorithm is to minimize the distance between the measured document and cluster centres using Eq. (2).

$$D_m = \sum_{i \in N} \sum_{j \in C} u_{ij}^m \cdot (d_{ik})^2 \quad , \qquad 1 \leq m \leq \infty \tag{2}$$

where

$$d_{ik} = \left[\sum_{j=1}^{m} (x_{kj} - v_{ij})^2\right]^{1/2}, \forall\, i \in N, k \in C \tag{3}$$

and x_{kj} is the document measured in the ith dimension, v_{ij} is the cluster vector in the ith dimension, m is the fuzziness exponent which is any real number greater than 1, u_{ij} is the membership function of TF-IDF weight of a given word, N is the total number of documents and C is the total number of clusters. The operation of the TIDC algorithm involves seven steps and the details are shown below:

Step 1: Initialize the fuzzy partition matrix, U = $[u_{ij}]$, $U^{(0)}$, in which $[u_{ij}] \in \mathbb{R}^{C \times N}$. In this step, the retrieved document firstly initialize the fuzzy partition matrix and assigns a membership function, associated with multiple clusters, ranging from 0 to 1.

Step 2: Distinguish the keywords for classifying web documents into meaningful categories.

Step 3: Calculate the TF-IDF weight of keywords in documents using Eq. (4). Two parts are involved in the TF-IDF weight, w_{tf}, which are term frequency (TF) part and inverse document frequency (IDF) part.

$$wt_{i,j} = tf_{i,j} \cdot idf_i \tag{4}$$

where $tf_{i,j}$ is the occurrence frequency of term i in document j and $idf_i = \log(\frac{N}{df_i})$ is the percentage of documents that contains the given word. N is the total number of web documents

obtained and df_i is the number of documents that contains the term i. Higher TF-IDF implies the more importance of the word.

Step 4: Compute the centre's vectors, $V^{(k)} = [v_{ij}]$, in which $[v_{ij}] \in \mathbb{R}^{C \times N}$, with U using Eq. (5).

$$v_{ij} = \frac{\sum_{i \in N} u_{ij}^{m} \cdot x_i}{\sum_{i \in N} u_{ij}^{m}} \tag{5}$$

Step 5: Update $U^{(k)}$, $U^{(k+1)}$ using Eq. (6).

$$u_{ij} = \left[\sum_{k \in N} \left(\frac{\|x_i - v_j\|}{\|x_j - v_k\|} \right)^{\frac{2}{m-1}} \right]^{-1} \tag{6}$$

Step 6: Exit the algorithm when $\| U^{(k+1)} - U^{(k)} \|$ is less than the stopping value between 0 to 1; otherwise repeat steps 4 to 5.

Step 7: Calculate the relevant value (R) of the individual web documents towards the inputted keywords. Therefore, web documents can be assigned into corresponding categories according to the membership function of the keywords. Moreover, the particulr value is adopted to indicate the relevance of the web documents towards the inputted keywords using Eq. (7).

$$R = \frac{\sum_{i \in V} w_{Ri} \cdot u_{ij}}{\sum_{i \in V} w_{Ri}} \tag{7}$$

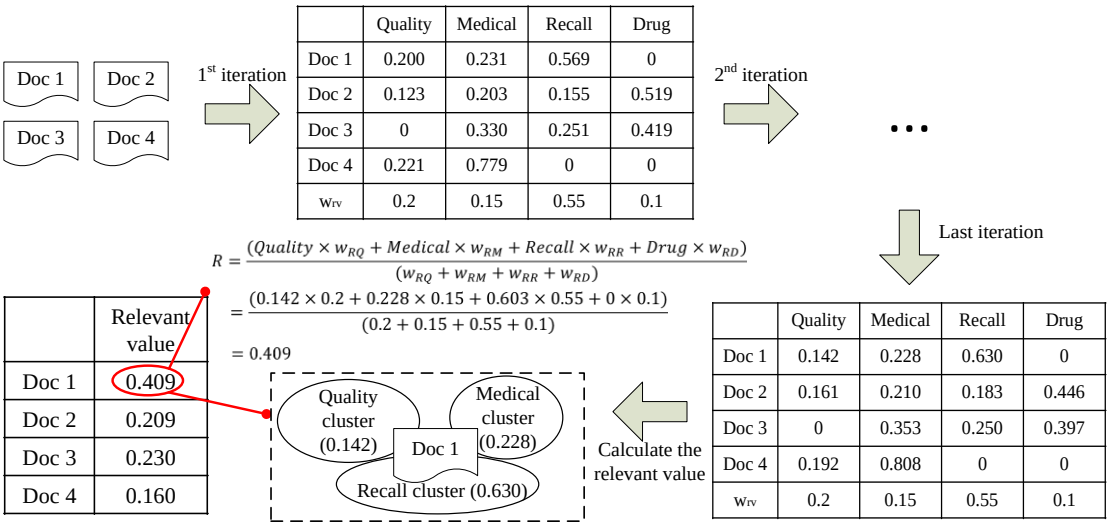
where w_R is the importance weight of the web documents to particular clusters. Figure 4 shows the mechanism for clustering and ranking the documents in the TIDC algorithm. By doing so, the WMQAS can automatically rank the extracted web documents in descending order according to their relevant value. Thus, users can easily search and discover the most important information.

To further reduce the time spent on reading the web documents by users, a short summary is generated for each document. By adding up the TF-IDF weight of terms in a sentence, the sentence with higher TF-IDF score indicates the higher importance of that sentence. Eq. (8) shows the formula for calculating the TF-IDF score of each sentence, s.

$$s = \sum_{i=1}^{k} wt_i \tag{8}$$

where s is the total score of the TF-IDF weight of the sentence for term i in sentence s. By extracting the sentence with the higher score and arranging into a short summary, users can easily understand the main idea of the web documents and hence extract the relevant information for revising the QA scheme. Finally, a new QA scheme with the appropriate

working instructions, identified incidents and recommendations can be formulated. The new QA scheme is then stored and retained in the case library as a reference for solving the new problem.



	Quality	Medical	Recall	Drug
Doc 1	0.200	0.231	0.569	0
Doc 2	0.123	0.203	0.155	0.519
Doc 3	0	0.330	0.251	0.419
Doc 4	0.221	0.779	0	0
w_{rv}	0.2	0.15	0.55	0.1

	Relevant value
Doc 1	0.409
Doc 2	0.209
Doc 3	0.230
Doc 4	0.160

	Quality	Medical	Recall	Drug
Doc 1	0.142	0.228	0.630	0
Doc 2	0.161	0.210	0.183	0.446
Doc 3	0	0.353	0.250	0.397
Doc 4	0.192	0.808	0	0
w_{rv}	0.2	0.15	0.55	0.1

Figure 4. The mechanism for clustering and ranking the documents in TIDC algorithm

4. Case Study

In order to validate the effectiveness of WMQAS, a pilot test was conducted in a Hong Kong-based logistics service provider, specializing in distribution operations, distribution logistics and value-adding services for multinational and local brand owners. It was founded in 1987 and owns a fleet of vehicles and pharmaceutical warehouses, certified under GDP and GSP, for distributing and storing pharmaceutical products for clients that includes homes for the elderly, clinics and pharmacies. Drugs, secondary packaging materials, consumable pharmaceutical products are common inventories stored in the pharmaceutical warehouse. Figures 5 and 6 show the secondary packaging materials stored in the warehouse and the storage environment for pharmaceutical products respectively. The company received GMP certification for carrying out the additional secondary packaging process in its own warehouse in recent years. Currently, the company relies on staff in the quality department to execute QA processes so as to ensure product quality. However, due to the complicated steps involved and the numerous elements to be considered in the GMP, it is hard for staff to execute such QA processes manually. Once mistakes and human error occur in the warehousing and packaging operations, the pharmaceutical products may be contaminated easily and hence result in high product rejection and recall rates. In addition, staff are required to spend long period of for manually reviewing medical and quality information every three months which is time consuming and ineffective. In order to improve the effectiveness of the QA scheme and maintain the quality of pharmaceutical products, the proposed WMQAS is implemented in the case company to facilitate decision making in formulating the work appropriate QA scheme

and identify the possible incidents in warehousing and packaging operations.

Figure 5. Secondary packaging materials stored in the warehouse

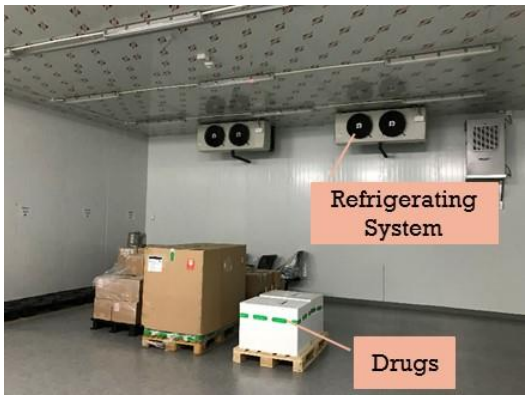


Figure 6. Storage environment for pharmaceutical products

4.1 Implementation of WMQAS

Figure 7 presents the implementation flow of the WMQAS which consists of five steps: (i) investigation of existing standards and practices in pharmaceutical warehouses, (ii) construction of the CBR library, (iii) case retrieval and ranking, (iv) case adaptation using the web mining technique and (v) case retention and storage.

4.1.1 Investigation of existing standards and practices in pharmaceutical warehouses

In order to understand the handling requirements for controlling the quality of pharmaceutical products, it is important to study the existing international standards and warehousing and packaging practices in the case company regularly. According to the requirements of GMP, staff should be aware of those working procedures and possible incidents that can affect the quality of pharmaceutical products in the current warehousing and packaging practice. They can make use of these information in formulating the QA scheme by the help of the WMQAS.

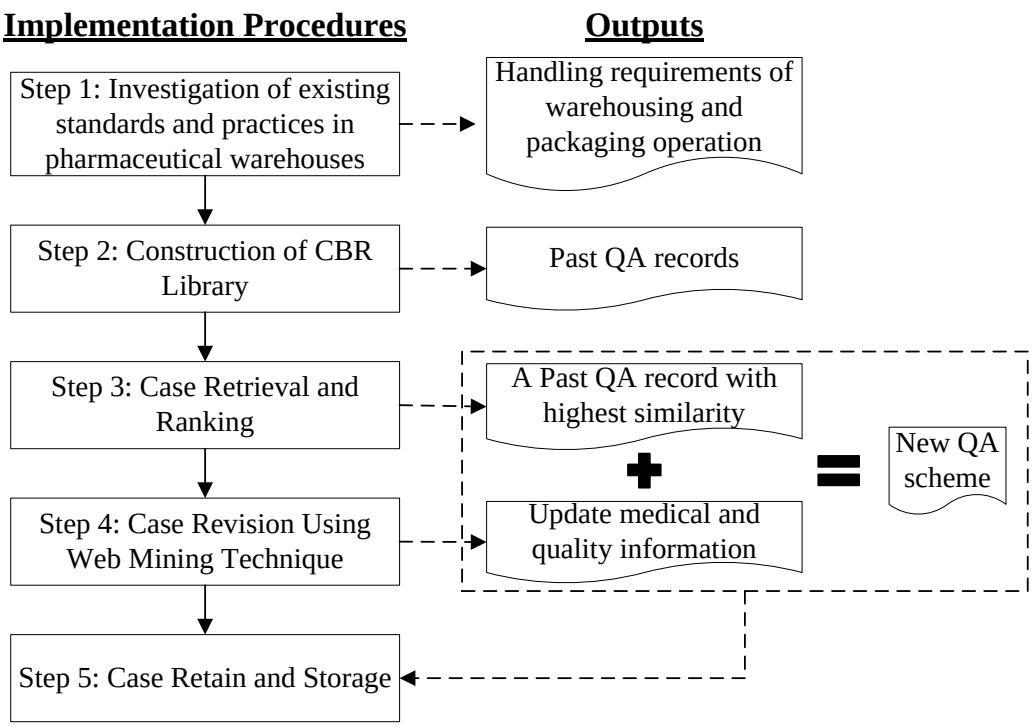


Figure 7. Implementation flow of the WMQAS

4.1.2 Construction of CBR library in WMQAS

In this phase, past QA records are stored as past reference cases for constructing the CBR library in WMQAS. The inductive indexing method is adopted to group past QA records into different clusters. A decision tree is constructed for retrieving the possible QA records which contains the key attributes in implementing the QA processes. These key attributes are predefined by experts and managers in the quality department according to the GMP requirements. Five levels are involved in the decision tree: (i) product categories, (ii) product type, (iii) temperature requirement, (iv) humidity requirement and (v) regulation compliance. Figure 8 shows the structure of the decision tree. The first level in the decision tree implies the most important attributes for clustering the past QA records in the CBR library. When passing along the search path in the decision tree, the number of cases with the corresponding attributes are narrowed down. As a result, the cases in the last level in the CBR library contain all the required attributes in the decision tree.

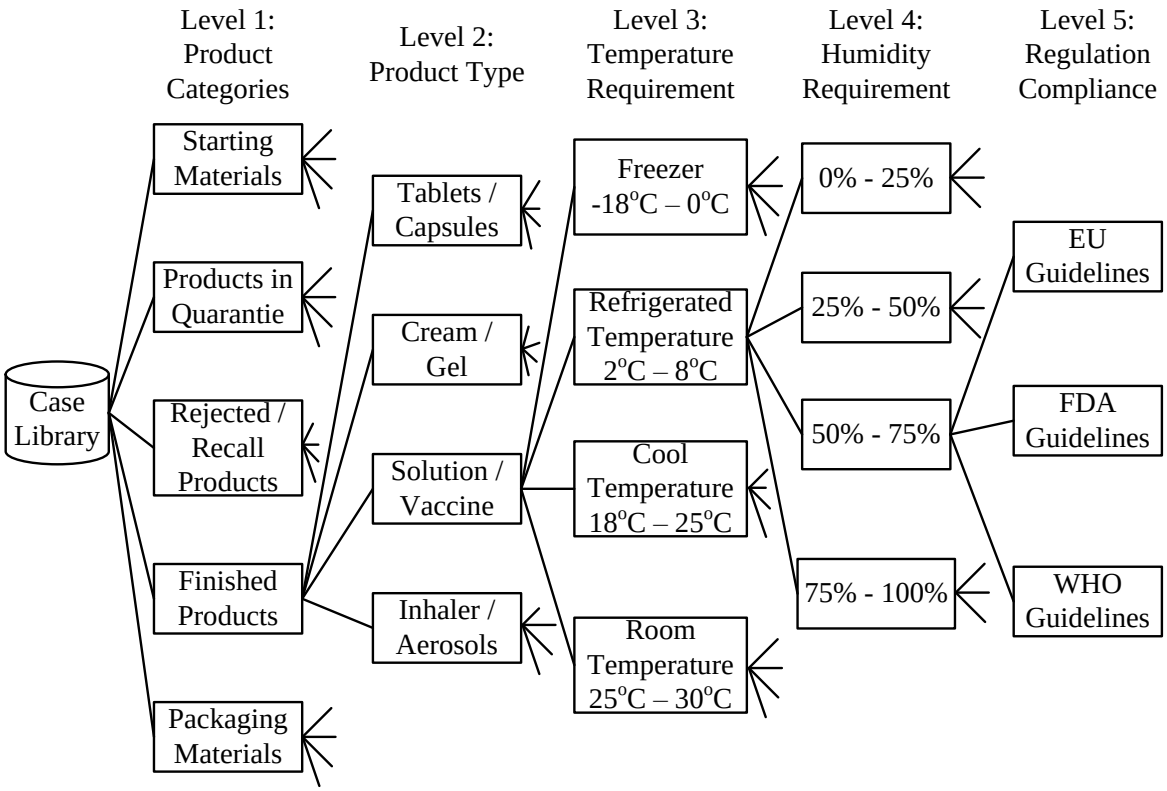


Figure 8. Structure of the decision tree

4.1.3 Case retrieval and ranking

When implementing the QA process in the pharmaceutical warehouse of the case company, the responsible staff has to input the required parameter values into the WMQAS. Based on the inputted value, the WMQAS retrieves the relevant group of QA records from the CBR library along the searching paths of the decision tree. The nearest neighbor method is then applied to rank the retrieved QA records, based on their similarity value using Eq. (1), which is shown in Section 3.2.1. According to the assessment requirements listed in GMP, several types of attributes or features, including secondary packaging services, packaging and labelling materials, and premises and equipment, were selected and weighted for calculating the similarity value of past QA records. Examples of attributes or features and the weightings are shown in Table 1. By summing up the similarity value of individual attributes, the total similarity values of the retrieved group of QA records can be calculated and then ranked in the descending order. Figure 9 shows the case retrieval and ranking process in the WMQAS. Based on the result in the case retrieval and ranking process, the QA staff can select the QA record with the highest similarity value for formulating a new QA plan in tackling the identified quality problems.

Table 1 Examples of attributes / features for calculating the similarity value

Attributes / Features (f)	Weighting (w_i)	Type of Attributes
Secondary Packaging Services		
- Cartoning	0.4	Textual
- Inserting Instruction Manuals	0.7	Textual
- Labeling	0.8	Textual
- Blister carding	0.4	Textual
Packaging and Labeling Materials		
- Name	0.8	Textual
- Expiry Date	1	Numeric
- Dosage	0.9	Numeric
- Description / Ingredients	0.8	Textual
Premises and Equipment		
- Clearness of Packaging Facilities	0.6	Textual
- Light Intensity	0.7	Numeric
- Carton Breakage Extent	0.4	Textual
- Maintenance of Equipment	0.5	Numeric
Others		
- Correct Clothing of Staff	0.7	Textual
- Correct Documentation Procedure	0.8	Textual
- Correct In-process Control	0.9	Textual

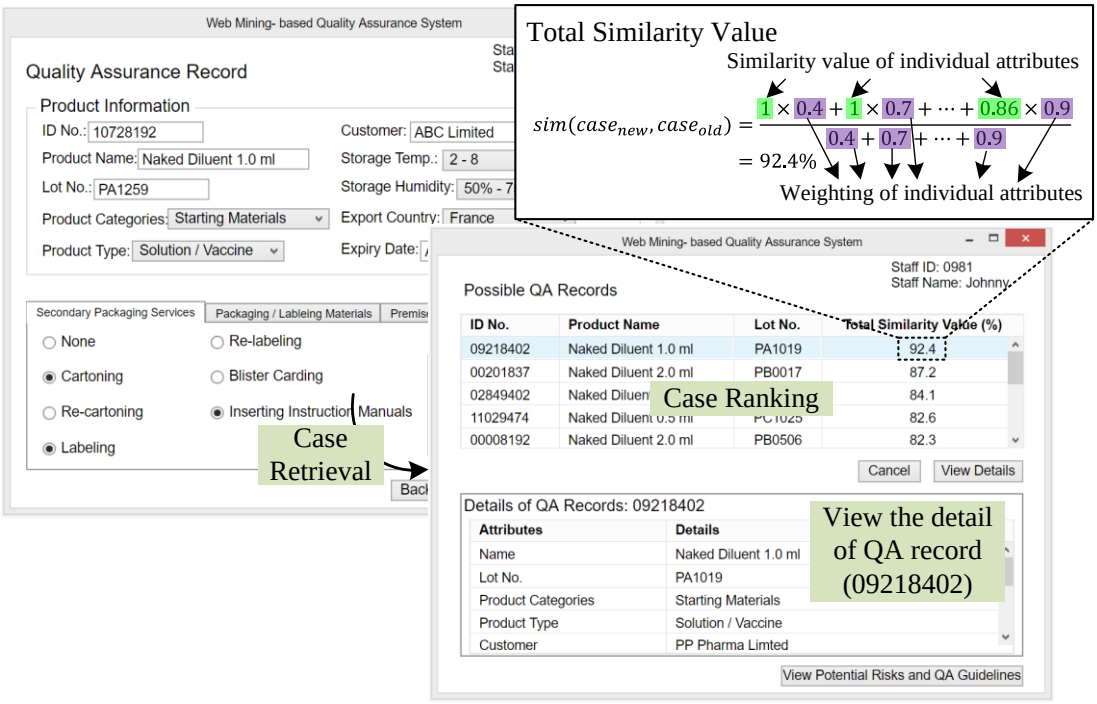


Figure 9. Case retrieval and ranking process

4.1.4 Case adaptation using web mining technique

Due to the continuous updating of products recall and safety alerts launched in the market,

the solution in the retrieved record may not be applicable for solving the current problem in the pharmaceutical warehouse. Therefore, in order to effectively extract updated information in handling the quality issues in the pharmaceutical warehouse, the TIDC algorithm is then adopted to the case adaptation process of CBR. Initially, QA staff have to identify and input the databases and websites that they visited frequently, such as “ECA Academy”, “Modern Healthcare”, “PubMed” and “FDA”, into the WMQAS. When they use the WMQAS to search for relevant information, it will automatically connect to these databases and websites which can significant improve the searching efficiency.

Using the TIDC algorithm, news and documents from the Internet are grouped into different categories, including “Quality”, “Medical” and “Operations”, based on the inputted keywords. “Quality” relates to actions that affect the quality of pharmaceutical products, including international guidelines and standards for monitoring and ensuring the pharmaceutical product quality. “Medical” refers to a brief description of the drugs which includes the ingredients and function of particular drugs, while “Operation” refers to the warehousing and packaging processes for handling the pharmaceutical products. The extracted relevant news and documents are then ranked in descending order according to the relevant value.

A case scenario is used to demonstrate how the TIDC algorithm can facilitate the QA staff to search for updated relevant information from the Internet. Seven steps are involved in the TIDC algorithm and details are shown as follows:

Step 1: Set the initial fuzzy partition matrix, $U^{(0)}$. In this example, 6 news and documents are selected for demonstrating the calculation of their membership value so as to group them into three predefined categories (c = 3). Assuming that the fuzziness exponent, m, is 2, the initial fuzzy partition matrix is

$$U^{(0)} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

This means that Doc 1 and Doc 4 belong to cluster 1, Doc 2 and Doc 5 belong to cluster 2 and Doc 3 and 6 belong to cluster 3 initially.

Step 2: Identify the keywords for searching the web information. In this case, QA staff have to find out the news and documents related to the recall of pharmaceutical products so as to increase their awareness in implementing the QA scheme. Three keywords “recall”, “process” and “pharmaceutical” are identified and entered in the WMQAS for searching the news and documents related to the recall of pharmaceutical products.

Step 3: Calculate the TF-IDF weight using Eq. (4). Figure 10 illustrates the calculation of the TF-IDF weight. The occurrence frequency of “recall”, “process” and “pharmaceutical” is the term frequency part in Eq. (4), as highlighted in Figure 9. For example, the word "recall" is counted as 46 in the news “Q1 2017 Pharma recall trends” which means that *tf* of “recall” is equal to 46 times. For inverse document frequency, 1036 news and documents (N = 1036) are stored in the WMQAS At the same time, there are 345 news and documents ($n_{(\text{recall})}$) that contain the keyword “recall”, and therefore, the value of *idf* is 0.478. By multiplying the *tf* and *idf* of keyword “recall”, the TF-IDF of keyword “recall” in the news “Q1 2017 Pharma recall trends” is found to be is 21.967. By doing so, the TF-IDF weight for the keywords “process” and “pharmaceutical” are calculated as 0.750 and 3.819, respectively. Table 2 shows the TF-IDF weighting of three inputted keywords in the six selected news and documents.

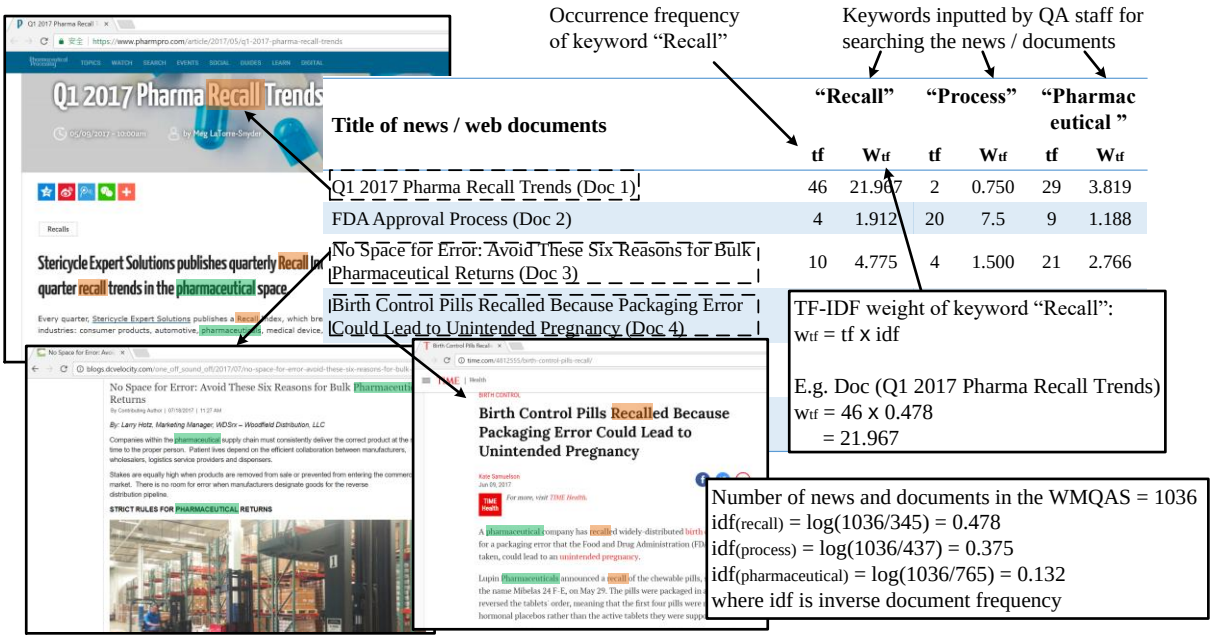


Figure 10. An example for calculating the TF-IDF weight of keywords

Table 2. TF-IDF weighting of three inputted keywords in the six selected news and documents.

	Doc 1	Doc 2	Doc 3	Doc 4	Doc 5	Doc 6
Recall	21.967	1.912	4.775	2.865	3.824	3.824
Process	0.75	7.5	1.5	1.974	1.5	5.625
Pharmaceutical	3.819	1.188	2.766	0.658	0.528	1.716

Step 4: Calculate the cluster vectors (v_{ij}) using Eq. (5). Take the quality cluster (C1) as an example, the membership value of Doc 1 and Doc 4 is 1 while the membership value of Doc 2, Doc 3, Doc 5 and Doc 6 is 0. Therefore, the vector of C1 is $v_1 = \{12.416, 1.362, 2.238\}$, C2 is v_2 = {2.868, 4.5, 0.858} and C3 is v_3 = {4.299, 3.563, 2.241}.

$$v_{1j} = \frac{u_{1j}^2 x_{1j} + u_{2j}^2 x_{2j} + u_{3j}^2 x_{3j} + u_{4j}^2 x_{4j} + u_{5j}^2 x_{5j} + u_{6j}^2 x_{6j}}{u_{1j}^2 + u_{2j}^2 + u_{3j}^2 + u_{4j}^2 + u_{5j}^2 + u_{6j}^2}$$

$$= \frac{(1)x_{1j} + (0)x_{2j} + (0)x_{3j} + (1)x_{4j} + (0)x_{5j} + (0)x_{6j}}{1 + 0 + 0 + 1 + 0 + 0}$$

$$\left.\begin{aligned} v_{11} &= \frac{21.967 + 2.865}{2} \\ v_{12} &= \frac{1.912 + 3.824}{2} \\ v_{13} &= \frac{4.775 + 3.824}{2} \end{aligned}\right\} v_1 = \{12.416, 1.362, 2.238\}$$

With the centers of clusters, the distance between the documents and clusters (d_{ik}) can be found using Eq. (3). For example, the distance between Doc1 and C1 is $d_{11} = \sqrt{(21.967 - 12.416)^2 + (0.75 - 1.362)^2 + (3.819 - 2.2385)^2} = 9.700$. This step is repeated for the remaining documents and cluster. The result of the distance measure is shown in Table 3.

Table 3. Distance of each documents from each cluster center

	Doc 1	**Doc 2**	**Doc 3**	**Doc 4**	**Doc 5**	**Doc 6**
Quality (C1)	9.700	12.211	7.660	9.700	8.762	9.606
Medical (C2)	19.688	3.169	4.034	2.534	3.166	1.708
Operation (C3)	17.959	4.724	2.181	2.662	2.723	2.181

Step 5: Update the fuzzy partition matrix ($U^{(1)}$) based on the distance of each document from each cluster center using Eq. (6). The new membership value of the six documents to C1 is shown as below.

$$u_{11} = \left[\sum_{j=1}^{c}\left(\frac{d_{11}}{d_{j1}}\right)^2\right]^{-1} = \left[\left(\frac{d_{11}}{d_{11}}\right)^2 + \left(\frac{d_{11}}{d_{21}}\right)^2 + \left(\frac{d_{11}}{d_{31}}\right)^2\right]^{-1} = \left[\left(\frac{9.700}{9.700}\right)^2 + \left(\frac{9.700}{19.688}\right)^2 + \left(\frac{9.700}{17.959}\right)^2\right]^{-1}$$

$$= 0.652$$

$$u_{12} = \left[\left(\frac{d_{12}}{d_{12}}\right)^2 + \left(\frac{d_{12}}{d_{22}}\right)^2 + \left(\frac{d_{12}}{d_{32}}\right)^2\right]^{-1} = \left[\left(\frac{12.211}{12.211}\right)^2 + \left(\frac{12.211}{3.169}\right)^2 + \left(\frac{12.211}{4.724}\right)^2\right]^{-1} = 0.044$$

$$u_{13} = \left[\left(\frac{d_{13}}{d_{13}}\right)^2 + \left(\frac{d_{13}}{d_{23}}\right)^2 + \left(\frac{d_{13}}{d_{33}}\right)^2\right]^{-1} = \left[\left(\frac{7.660}{7.660}\right)^2 + \left(\frac{7.660}{4.034}\right)^2 + \left(\frac{7.660}{2.181}\right)^2\right]^{-1} = 0.059$$

$$u_{14} = \left[\left(\frac{d_{14}}{d_{14}}\right)^2 + \left(\frac{d_{14}}{d_{24}}\right)^2 + \left(\frac{d_{14}}{d_{34}}\right)^2\right]^{-1} = \left[\left(\frac{9.700}{9.700}\right)^2 + \left(\frac{9.700}{2.534}\right)^2 + \left(\frac{9.700}{2.662}\right)^2\right]^{-1} = 0.035$$

$$u_{15} = \left[\left(\frac{d_{15}}{d_{15}}\right)^2 + \left(\frac{d_{15}}{d_{25}}\right)^2 + \left(\frac{d_{15}}{d_{35}}\right)^2\right]^{-1} = \left[\left(\frac{8.762}{8.762}\right)^2 + \left(\frac{8.762}{3.166}\right)^2 + \left(\frac{8.762}{2.723}\right)^2\right]^{-1} = 0.052$$

$$u_{16} = \left[\left(\frac{d_{16}}{d_{16}}\right)^2 + \left(\frac{d_{16}}{d_{26}}\right)^2 + \left(\frac{d_{16}}{d_{36}}\right)^2\right]^{-1} = \left[\left(\frac{9.606}{9.606}\right)^2 + \left(\frac{9.606}{1.708}\right)^2 + \left(\frac{9.606}{2.181}\right)^2\right]^{-1} = 0.019$$

By doing so, the remaining fuzzy partition values are calculated and the updated fuzzy partition matrix is

$$U^{(1)} = \begin{bmatrix} 0.652 & 0.044 & 0.059 & 0.035 & 0.052 & 0.019 \\ 0.158 & 0.660 & 0.213 & 0.506 & 0.403 & 0.608 \\ 0.190 & 0.296 & 0.728 & 0.459 & 0.545 & 0.373 \end{bmatrix}$$

Step 6: Check the difference between the updated and original membership value. Exit the FCM algorithm if $\|U^{(k+1)} - U^{(k)}\|$ is less than the stopping condition, i.e. the threshold value is smaller than 0.01. Since all the changes of the membership value are greater than 0.01, the second iteration is required which defines $k = k + 1$ and then return to Step 4.

After six iteration, all the changes of the membership values meet the stopping condition and the FCM algorithm is exited. The final fuzzy partition matrix in this case is

$$U^{(6)} = \begin{bmatrix} 0.999^{*} & 0.004 & 0.010 & 0.003 & 0.002 & 0.005 \\ 0.000 & 0.953^{*} & 0.094 & 0.055 & 0.019 & 0.881^{*} \\ 0.001 & 0.043 & 0.896^{*} & 0.942^{*} & 0.979^{*} & 0.114 \end{bmatrix}$$

where * is the highest fuzzy membership value of the document to the particular cluster. From the result of the FCM algorithm, it shows that Doc 1 strongly belongs to the Quality cluster, Doc 2 and Doc 6 strongly belong to the Medical cluster, and Doc 3, Doc 4 and Doc 5 belong to the Operation cluster.

Step 7: Compute the relevant value (R) of each document with reference to the importance of each cluster based on the membership value of the documents to each cluster and Eq. (7). Since the QA staff have to search for the recall information of pharmaceutical products which may affect the product quality, the weighting of quality category is ranked as the most important. Moreover, the ranking of Medical and Operation categories is important and relevant. Normalization is then used to present the relevant value of the news and documents, ranging from 0 to 1, using Eq. (9).

$$\tilde{R} = \frac{R}{R_o} \tag{9}$$

where $R_o = \max\left\{w_{Ri} \times u_{ij}\right\}, \forall\, i, j \in R$ and R is the relevant value of an individual document. Figure 11 shows the search results of WMQAS based on the three inputted keywords. Since the news "Q1 2017 Pharma Recall Trend" is the most relevant news with a relevant value 1, QA staff can read a short summary of the news to capture the desirable information for modifying the past QA records. From that news, it is found that failed specifications and mislabeling issues are the top two reasons that caused the recall of pharmaceutical products in the first quarter of 2017. Therefore, in implementing the QA scheme, QA staff should add these two possible incidents in the retrieved QA records so as to increase their awareness for maintaining the high product quality. Figure 12 presents the

transformation of the useful web information, i.e. the top two reasons that caused the recall of pharmaceutical products, in modifying the retrieved QA record.

4.1.5 Case retention and storage

The revised QA record with update medical and quality information is treated as the new QA scheme. This new QA scheme can facilitate the decision making of QA staff so as to identify the possible incidents in implementing the QA process and hence provide solutions for mitigating the identified incidents so as to ensure the product quality. With the increasing number of cases in the CBR library, the quality of QA scheme generated by the WMQAS can be improved due to the learning ability of the CBR engine.

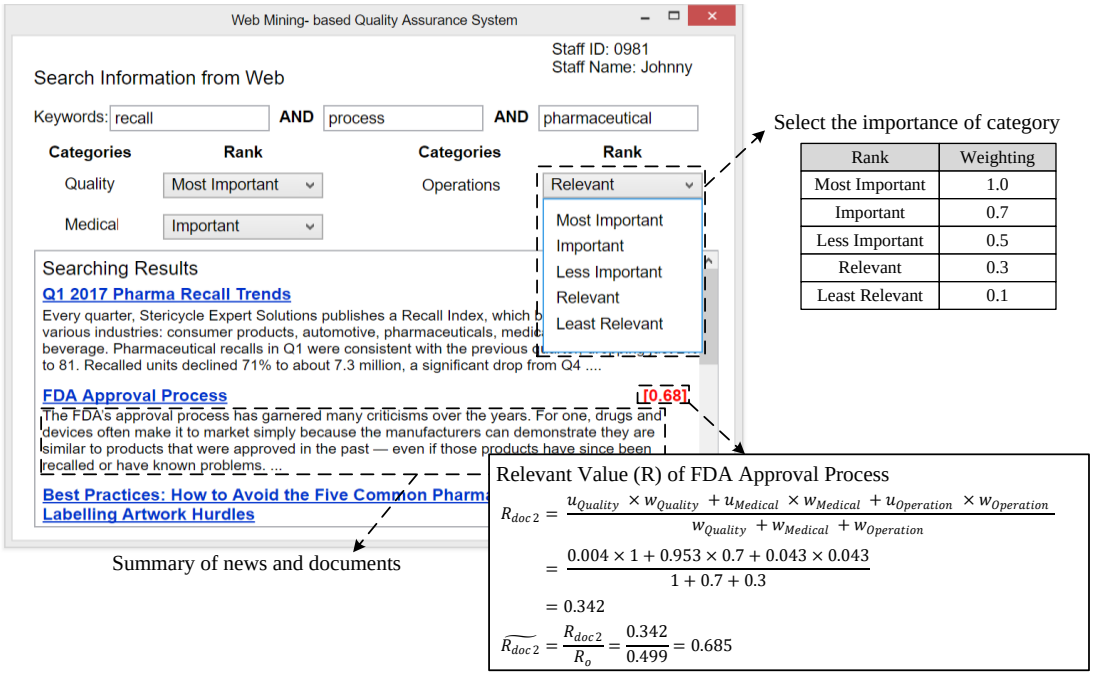


Figure 11. The search results of WMQAS

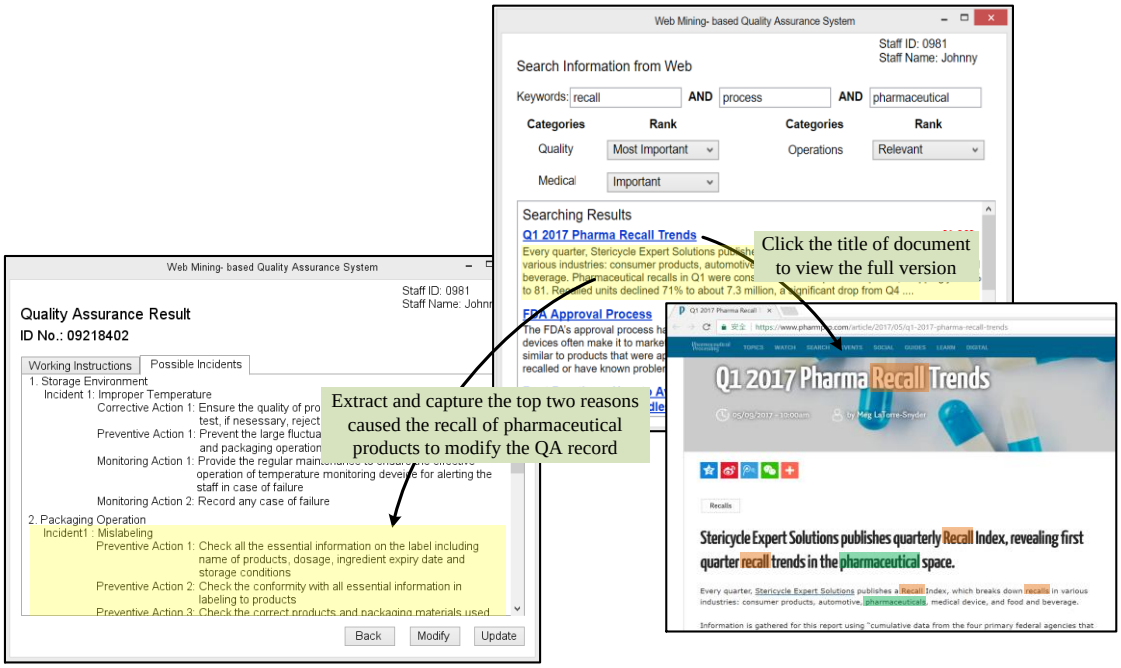


Figure 12. Transformation of extracted information for modifying the QA record

5. Results and Discussion

A pilot study was conducted in the case company to validate the feasibility of the proposed WMQAS. During a six-month trial run in the case company, the QA operation process for controlling the pharmaceutical products' quality was enhanced. In this section, the advantages and performance of the WMQAS are discussed. The managerial implication to the LSPs is also mentioned.

5.1 Advantages of the WMQAS in the case company

There are two advantages of the WMQAS: (i) improvement in QA scheme formulation and (ii) improvement in product quality and customer satisfaction. Detailed discussion is as follows.

5.1.1 Improvement in QA scheme formulation

Instead of the manual QA scheme formulation, a systematic approach is used in decision making process by the application of WMQAS. Four criteria are identified for assessing the QA scheme formulation efficiency: (i) time for providing the working instructions, (ii) time for identifying the possible incidents in pharmaceutical warehouses, (iii) time for searching quality and medical information and (iv) time for suggesting solutions to mitigate the incidents. Table 4 shows the improvement in QA scheme formulation. Initially, QA staff rely on the personal experience and subjective judgment to execute the complicated QA processes. In addition, they require to spend a lot of time to review the web information so as to update their knowledge in handling the pharmaceutical products. With the use of the CBR engine, the time for providing

the working instructions for carrying the warehousing and packaging operations can be significantly reduced from 20 minutes to 5 minutes, which has 75.0% improvement. In addition, QA staff can effectively identify the possible incidents in and generate solutions by making reference to past QA records instead of by manual manipulation. Therefore, the time required for these steps are reduced by 64.7% and 66.7%, respectively. Moreover, compared to the traditional method in finding the relevant information from different websites and databases, the time for searching for quality and medical information was significantly reduced by 90.6%. This is mainly because the WMQAS allows QA staff to search for quality and medical information by inputting the keywords in a single platform. As a result, the total time for formulating the QA scheme is significantly reduced from 140 minutes to 25 minutes, an 82.14% improvement. Apart from the improvement of efficiency in formulating the QA scheme, WMQAS also improves its effectiveness. In this case, effectiveness means that the QA scheme generated by the QA staff can directly generate the appropriate working instructions and mitigate identified incidents in the pharmaceutical warehouse without modification. Without the WMQAS, the effectiveness of the QA scheme is only 45% while the effectiveness of QA scheme is improved to 86% when using the WMQAS. Since the solutions from past QA records can be taken as reference for solving the current situation when handling new pharmaceutical products, it is found that the effectiveness of QA scheme is enhanced by 47.7%.

Table 4. Improvement in QA scheme formulation

	Without WMQAS	With WMQAS	Improvement (%)
Time for formulating the QA scheme			
- Provide working instructions	20 min	5 min	75.0%
- Identify the possible incidents	17 min	6 min	64.7%
- Search quality and medical information	85 min	8 min	90.6%
- Suggest solutions for mitigating the incidents	18 min	6 min	66.7%
Total	**140 min**	**25 min**	**82.14%**
Effectiveness of QA scheme	45%	86%	47.7%

5.1.2 Improvement in the product quality and customers satisfaction

Without the WMQAS, quality assurance is carried out by the QA staff based on their experience and knowledge. Any human error or mistake that occurred in the decision making process may be hided the incidents in pharmaceutical products. This may lead to the recall and rejection of pharmaceutical products due to reasons of contamination, failure specification and mislabeling issues. With the help of the WMQAS, possible incidents that may occur in warehousing and packaging can be easily identified by retrieving past QA records as reference, while an effective QA plan with proper working instructions is generated automatically for preventing the incidents occurred. Thus, errors and mistakes are significantly reduced in

implementing the QA scheme. In addition, since the historical QA records are stored in the case library of WMQAS, the QA staff can easily keep track of these documents for ensuring that the quality assurance process complies with the required standards, such as GMP and FDA. Table 5 presents the improvement in product quality and customer satisfaction. It is found that the number of recalls and rejections per six months is reduced from 6 to 2. In addition, with the decrease in the number of recalls and rejections of pharmaceutical products, the average company loss due to the poor product quality over six months is reduced by 70.6% while the customer satisfaction, i.e. homes for the elderly, clinics and pharmacies, is increased by 27.7%.

Table 5. Improvement in the product quality and customers satisfaction

	Without WMQAS	With WMQAS	Improvement (%)
Recall and rejection of pharmaceutical products (six months)	6	2	66.7%
Average loss due to the poor product quality (in thousand)	635	187	70.6%
Customer satisfaction (%)	65.7%	83.9%	27.7%

5.2 Discussion on the performance of WMQAS

Since news related to the recall and the modification of quality standards for handling pharmaceutical products may be frequently revised, the TIDC algorithm allows QA staff to effective capture such information by entering the keywords in the WMQAS. Compared with the adoption of the TIDC algorithm in the case adaption process of CBR, in traditional CBR, staff are required to modify the retrieved QA records based on their experience. This may affect the consistency of the QA plan for handling the risks for the same pharmaceutical products. In order to show that the TIDC algorithm can improve the consistency of the plan, an experiment was carried out to compare the results of the QA plan using traditional CBR and the WMQAS. QA staff with different working experience is selected and divided into two groups to use the two approaches for generating the QA plan for mitigating risks. Group A involved senior staff with five years or more than five years working experience while Group B involved junior staff with less than five years working experience, respectively. The findings of the experiment are as follows:

(i) The QA scheme generated by both approaches, i.e. traditional CBR and proposed WMQAS, helps to mitigate the risks in pharmaceutical warehouses and hence maintains products quality.

(ii) For group A, variation of the QA solution using the proposed WMQAS is 5% which is slightly lower than the variation of QA solution using traditional CBR (8%). For group B, there is a significant decrease in the variation of the QA solution when using proposed WMQAS (from 46% to 10%).

(iii) Variation of the QA solution within group A is significantly lower than group B in both approaches. It is reasonable that senior staff have more experience in identifying the risks and in formulating the QA scheme. For junior staff, they may fail to identify the possible risks in warehousing and packaging operations due to lack of experience.

(iv) With the use of the proposed WMQAS, the variation of the QA solution within group B is reduced to 10%. This implies that the WMQAS can overcome the lack of experience in junior staff so as to facilitate decision making in formulating a QA scheme with high quality.

5.3 Managerial implication

Considering the rapid global aging population, ensuring good quality of pharmaceutical products along the supply chain is important to maintain the designated level of medical efficacy and therapeutic effect. As one of major parties in the pharmaceutical supply chain, LSPs, such as distributors, are striving to establish appropriate QA with working instructions and procedures for carrying the additional secondary packaging and warehousing process in their own warehouses. However, in the absence of decision making approach facilitating the formulation of QA scheme in pharmaceutical warehouses, LSPs experience obstacles in maintaining the quality of such products due to its complicated handling characteristics including the high product variety, smaller product size, high similarity in product ingredients and environmental sensitivity. This may result in product recall and poor customer satisfaction if warehousing procedures are inappropriately conducted.

As indicated in the results of the study, the proposed system not only provides working guidance for implementation of QA scheme, but also identifies the possible incidents from operational point of view for seconding packaging and warehousing operations. It enhances the management in QA processes so as to reduce the errors and loss due to the rejection or recall of such products in the case company. In addition, the adoption of web mining helps fill the knowledge gap among various level of staff. The findings make a contribution to LSPs for increasing their awareness, knowledge and capability to address the quality issues in pharmaceutical supply chain. LSPs are able to enrich the knowledge in handling pharmaceutical products by accurately capturing update medical and quality from international institutions and official electronic documents. In broader view, with such information, LSPs are able to be easily aware of the trend of the pharmaceutical industry and hence new services for fulfilling the customers' needs can be created.

6. Conclusions

With the increasing demand for pharmaceutical products, ensuring high quality of such products is critical to protect the health of patients. Due to the characteristics of such products, which are smaller in size, with wide variety and highly sensitive to the change of storage

environment, an effective quality assurance scheme is critical for LSPs in handling such products in warehousing and packaging operations when complying with the GMP. This paper described a web mining-based quality assurance system (WMQAS) to facilitate the decision making of staff in formulating an effective QA scheme in pharmaceutical warehouses. With the use of the CBR technique, appropriate working instructions and procedures can be generated and a list of possible incidents that may occur in warehousing and packaging operations can be identified based on the past QA records. This can reduce the human effort in carrying out the complicated steps involved in the GMP. Apart from making use of human knowledge for revising the retrieved past QA record, online knowledge acquisition approach is also adopted in the case adaptation process of CBR. The TIDC algorithm embedded in the case adaptation process allows QA staff to cluster vast information into meaningful categories and hence to obtain the most relevant web information efficiently. By extracting web information from the Internet, the modification of QA records can be undertaken so as to improve the quality of the QA scheme in mitigating the new problems.

The proposed WMQAS was developed and implemented in the Hong Kong-based LSP. The results show that the WMQAS not only improves the effectiveness of QA processes, but also reduces the recall and rejection rates of pharmaceutical products. In addition, the WMQAS allows the QA staff, especially junior staff, to generate an appropriate QA scheme with good quality for examining products in pharmaceutical warehouses. However, there is a limitation in the current study. In the case company, three categories are used for classifying news and documents in the TIDC algorithm. The accuracy of the search result may be affected when using different numbers of clusters. Further research on optimizing the number of clusters for grouping web information is required for improvement. Besides, the stopping condition in the TIDC algorithm is defined as smaller than 0.01 in this paper and investigation on using different values of stopping condition is suggested. Even though the system is currently designed for pharmaceutical product warehousing systems, it could also be applied for enhancing quality assurance of other products with similar characteristics such as premium fresh fruit, meal, seafood and wine.

References

Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI communications*, *7*(1), 39-59.

Abrahams, A. S., Fan, W., Wang, G. A., Zhang, Z. J., & Jiao, J. (2015). An integrated text analytic framework for product defect discovery. *Production and Operations Management*, *24*(6), 975-990.

Al-Ayyoub, M., Abu-Dalo, A. M., Jararweh, Y., Jarrah, M., & Al Sa'd, M. (2015). A gpu-based implementations of the fuzzy c-means algorithms for medical image segmentation. *The*

Journal of Supercomputing, *71*(8), 3149-3162.

Alli, I. (2016). *Food quality assurance: principles and practices*. CRC Press.

Alptekin, G. I., & Büyüközkan, G. (2011). An integrated case-based reasoning and MCDM system for Web based tourism destination planning. *Expert Systems with Applications*, *38*(3), 2125-2132.

Anand, G., Gray, J., & Siemsen, E. (2012). Decay, shock, and renewal: operational routines and process entropy in the pharmaceutical industry. *Organization Science*, *23*(6), 1700-1716.

Asian Organization of Supreme Audit Institutions (2009) Quality Assurance in Financial Auditing, A Handbook, ASOSAI.

Aung, M. M., & Chang, Y. S. (2014). Temperature management for the quality assurance of a perishable food supply chain. *Food Control*, *40*, 198-207.

Belecheanu, R., Pawar, K. S., Barson, R. J., Bredehorst, B., & Weber, F. (2003). The application of case based reasoning to decision support in new product development. *Integrated Manufacturing Systems*, *14*(1), 36-45.

Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, *10*(2-3), 191-203.

Castro, J. L., Navarro, M., Sánchez, J. M., & Zurita, J. M. (2011). Introducing attribute risk for retrieval in case-based reasoning. *Knowledge-Based Systems*, *24*(2), 257-268.

Chakraborty, S., & Boral, S. (2017). A developed case-based reasoning system for machine tool selection. *Benchmarking: An International Journal*, *24*(5), 1364-1385.

Chebel-Morello, B., Haouchine, M. K., & Zerhouni, N. (2013). Reutilization of diagnostic cases by adaptation of knowledge models. *Engineering Applications of Artificial Intelligence*, *26*(10), 2559-2573.

Drug Office, (2017). Guidance for Industry: Hong Kong Guide to GMP for the Secondary Packaging of Pharmaceutical Products. Department of Health, Retrieved from http://www.drugoffice.gov.hk/eps/do/en/doc/guidelines_forms/guidance_for_Industry_eng.pdf?v=p7ev5

Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, *3*(3), 32-57.

El-Menhawy, A., Mohamed, N. M., Arafa, H. A., & Elmonem, A. A. (2015). Improving the Performance of Pharmaceutical Tablet Production using Six Sigma Methodology (Modulation on Digestive Enzyme Tablet). *International Journal of Pharmaceutical Sciences and Research*, *6*(8), 3580-3593.

Fan, Z. P., Li, Y. H., Wang, X., & Liu, Y. (2014). Hybrid similarity measure for case retrieval in CBR and its application to emergency response towards gas explosion. *Expert Systems with Applications*, *41*(5), 2526-2534.

Felice, F. D., Petrillo, A., & Autorino, C. (2014). Maintenance Strategies and Innovative

Approaches in the Pharmaceutical Industry: An Integrated Management System (IMS). *International Journal of Engineering Business Management*, *6*(29), 1-9.

Fuchs, B., Lieber, J., Mille, A., & Napoli, A. (2014). Differential adaptation: An operational approach to adaptation for solving numerical problems with CBR. *Knowledge-Based Systems*, *68*, 103-114.

Ghosh, S., & Dubey, S. K. (2013). Comparative analysis of k-means and fuzzy c-means algorithms. *International Journal of Advanced Computer Science and Applications*, *4*(4), 35-39.

Gray, J. V., Roth, A. V., & Leiblein, M. J. (2011). Quality risk in offshore manufacturing: Evidence from the pharmaceutical industry. *Journal of Operations Management*, *29*(7), 737-752.

Gu, J., Goetschalckx, M., & McGinnis, L. F. (2007). Research on warehouse operation: A comprehensive review. *European journal of operational research*, *177*(1), 1-21.

Haleem, R. M., Salem, M. Y., Fatahallah, F. A., & Abdelfattah, L. E. (2015). Quality in the pharmaceutical industry–A literature review. *Saudi Pharmaceutical Journal*, *23*(5), 463-469.

Haque, B. U., Belecheanu, R. A., Barson, R. J., & Pawar, K. S. (2000). Towards the application of case based reasoning to decision-making in concurrent product development (concurrent engineering). Knowledge-Based Systems, *13*(2), 101-112.

He, W., Xu, L. D., Means, T., & Wang, P. (2009). Integrating web 2.0 with the case-based reasoning cycle: A systems approach. Systems Research and Behavioral Science, *26*(6), 717-728.

He, W., & Tian, X. (2017). A longitudinal study of user queries and browsing requests in a case-based reasoning retrieval system. *Journal of the Association for Information Science and Technology*. *68*(5), 1124–1136.

He, W. (2011). Using text mining to uncover students' technology-related problems in live video streaming. *British Journal of Educational Technology*, *42*(1), 40-49.

Heinrich, M. (2015). Quality and safety of herbal medical products: regulation and the need for quality assurance along the value chains. *British journal of clinical pharmacology*, *80*(1), 62-66.

Jalali, V., & Leake, D. (2016). Enhancing case-based regression with automatically-generated ensembles of adaptations. *Journal of Intelligent Information Systems*, *46*(2), 237-258.

Jie, F., Parton, K. A., & Mustafid. (2016). Supply chain performance flexibility in the Australian beef industry. *International Journal of Logistics Research and Applications*, *19*(4), 300-317.

Kanathey, K., Thakur, R. S., & Jaloree, S. (2018). Ranking of Web Pages Using Aggregation of Page Rank and HITS Algorithm. *International Journal of Advanced Studies in*

Computers, Science and Engineering, *7*(2), 17-22.

Karapetrovic, S., & Willborn, W. (2000). Quality assurance and effectiveness of audit systems. *International Journal of Quality & Reliability Management*, *17*(6), 679-703.

Kim, J., Han, M., Lee, Y., & Park, Y. (2016). Futuristic data-driven scenario building: Incorporating text mining and fuzzy association rule mining into fuzzy cognitive map. *Expert Systems with Applications*, *57*, 311-323.

Kolodner J. (1993). *Case-based reasoning*. Morgan Kaufmann, San Mateo

Kumar, N., & Jha, A. (2015). Latest trend in drugs regulatory guidance on 'good manufacturing practices'. *International Journal of Pharmaceutical Sciences and Business Management*, *3*(10), 10-16

Lam, H. Y., Choy, K. L., Ho, G. T. S., Kwong, C. K., & Lee, C. K. M. (2013). A real-time risk control and monitoring system for incident handling in wine storage. *Expert systems with applications*, *40*(9), 3665-3678.

Lao, S. I., Choy, K. L., Ho, G. T. S., Yam, R. C., Tsim, Y. C., & Poon, T. C. (2012). Achieving quality assurance functionality in the food industry using a hybrid case-based reasoning and fuzzy logic approach. *Expert Systems with Applications*, *39*(5), 5251-5261.

Law, D., Gruss, R., & Abrahams, A. S. (2017). Automated defect discovery for dishwasher appliances from online consumer reviews. *Expert Systems with Applications*, *67*, 84-94.

Lawrence, X. Y., Amidon, G., Khan, M. A., Hoag, S. W., Polli, J., Raju, G. K., & Woodcock, J. (2014). Understanding pharmaceutical quality by design. *The AAPS journal*, *16*(4), 771-783.

Lee, C. K. H., Choy, K. L., & Chan, Y. N. (2014). A knowledge-based ingredient formulation system for chemical product development in the personal care industry. *Computers & Chemical Engineering*, *65*, 40-53.

Li, H., & Sun, J. (2011). Principal component case-based reasoning ensemble for business failure prediction. *Information & Management*, *48*(6), 220-227.

Lim, M. K., Bahr, W., & Leung, S. C. (2013). RFID in the warehouse: A literature analysis (1995–2010) of its applications, benefits, challenges and future trends. *International Journal of Production Economics*, *145*(1), 409-430.

Mahzan, N., & Hassan, N. A. B. (2015). Internal Audit of Quality in 5s Environment: Perception on Critical Factors, Effectiveness and Impact on Organizational Performance. *International Journal of Academic Research in Accounting, Finance and Management Sciences*, *5*(1), 92-102.

Main, J., Dillon, T. S., & Shiu, S. C. (2001). A tutorial on case based reasoning. In *Soft computing in case based reasoning* (pp. 1-28). Springer London.

Mangan, J., & Lalwani, C. (2016). *Global logistics and supply chain management*. John Wiley & Sons.

Marinkovic, V., Bekcic, S., Pejovic, G., Sibalija, T., Majstorovic, V., & Tasic, L. (2016). An

approach to TQM evaluation in pharma business. *The TQM Journal*, *28*(5), 745-759.

Min, H. (2010). Artificial intelligence in supply chain management: theory and applications. *International Journal of Logistics Research and Applications*, *13*(1), 13-39.

Moreno-Luzon, M. D., & Peris, F. J. (1998). Strategic approaches, organizational design and quality management: integration in a fit and contingency model. *International Journal of Quality Science*, *3*(4), 328-347.

Na, S., Xumin, L., & Yong, G. (2010, April). Research on k-means clustering algorithm: An improved k-means clustering algorithm. In *Intelligent Information Technology and Security Informatics (IITSI), 2010 Third International Symposium on* (pp. 63-67). IEEE.

Newton, P. N., Schellenberg, D., Ashley, E. A., Ravinetto, R., Green, M. D., ter Kuile, F. O., Tabernero, P., White, N. J., & Guerin, P. J. (2015). Quality assurance of drugs used in clinical trials: proposal for adapting guidelines. BMJ: British Medical Journal, 350, 1-3.

Oláh, J., Karmazin, G., Pető, K., & Popp, J. (2017). Information technology developments of logistics service providers in Hungary. *International Journal of Logistics Research and Applications*, 1-13.

Ozer, M. (2005). Fuzzy c-means clustering and Internet portals: A case study. *European Journal of Operational Research*, *164*(3), 696-714.

Paine, F. A., & Lockhart, H. (2012). Packaging of pharmaceuticals and healthcare products. Springer Science & Business Media, New York.

Papalexi, M., Bamford, D., & Dehe, B. (2016). A case study of kanban implementation within the pharmaceutical supply chain. *International Journal of Logistics Research and Applications*, *19*(4), 239-255.

Perçin, S., & Min, H. (2013). A hybrid quality function deployment and fuzzy decision-making methodology for the optimal selection of third-party logistics service providers. *International Journal of Logistics Research and Applications*, *16*(5), 380-397.

Pharmacy and Poisons Board of Hong Kong (2014). Hong Kong Guide to Good Manufacturing Practice for Secondary Packaging of Pharmaceutical Products, 1-19.

Qi, J., Hu, J., & Peng, Y. (2017). A modularized case adaptation method of case-based reasoning in parametric machinery design. *Engineering Applications of Artificial Intelligence*, *64*, 352-366.

Quinn, F. J. (2013) Clinical trial logistics and packaging, Retrieved from http://pharmaceuticalcommerce.com/manufacturing-and-packaging/clinical-trial-logistics-and-packaging/ (retrieved 05.10.17).

Rao, R. S., & Arora, J. (2017). A Survey on Methods used in Web Usage Mining. *International Research Journal of Engineering and Technology, 4*(5)*,* 2627-2631

Reyes, E. R., Negny, S., Robles, G. C., & Le Lann, J. M. (2015). Improvement of online adaptation knowledge acquisition and reuse in case-based reasoning: Application to process engineering design. *Engineering Applications of Artificial Intelligence*, *41*, 1-16.

Riesbeck, C. K., & Schank, R. C. (2013). *Inside case-based reasoning*. Psychology Press.

Robinson, S. (2017). *Automated inspection and quality assurance*. Routledge.

Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, *51*(1), 368-384.

Rossetti, C. L., Handfield, R., & Dooley, K. J. (2011). Forces, trends, and decisions in pharmaceutical supply chain management. *International Journal of Physical Distribution & Logistics Management*, *41*(6), 601-622.

Salloum, S. A., Al-Emran, M., Monem, A. A., & Shaalan, K. (2017). A survey of text mining in social media: facebook and twitter perspectives. *Advances in Science, Technology and Engineering Systems Journal, 2*(1), 127-133.

Seo, J., Han, S., Lee, S., & Kim, H. (2015). Computer vision techniques for construction safety and health monitoring. *Advanced Engineering Informatics*, *29*(2), 239-251.

Sharifian, R., Ebrahimi, S., & Bastani, P. (2016). How Radio Frequency Identification Improves Pharmaceutical Industry: A Comprehensive Review Literature. *Journal of Pharmaceutical Care*, *3*(1), 26-33.

Shibata, H., Yoshida, H., Izutsu, K. I., Yomota, C., Goda, Y., & Okuda, H. (2016). Scientific and regulatory approaches to confirm quality and improve patient perceptions of generic drug products in Japan. *AAPS Open*, *2*(1), 6.

Thomaz, G. M., Biz, A. A., Bettoni, E. M., Mendes-Filho, L., & Buhalis, D. (2017). Content mining framework in social media: A FIFA world cup 2014 case analysis. *Information & Management*, *54*(6), 786-801.

Ting, S. L., See-To, E. W., & Tse, Y. K. (2013). Web information retrieval for health professionals. *Journal of Medical Systems,* 37(3), 1-14.

Troup, G. M., & Georgakis, C. (2013). Process systems engineering tools in the pharmaceutical industry. *Computers & Chemical Engineering*, *51*, 157-171.

Velayutham, S. (2003). The Accounting Profession's code of Ethics: is it a code of Ethics or a Code of Quality Assurance?. *Critical Perspectives on Accounting*, *14*(4), 483-503.

Wang, C. S., Wu, T. H., & Fan, S. K. S. (2013). An enhanced case-based diagnosis mechanism for abnormal production administration. *International Journal of Production Research*, *51*(10), 3131-3142.

Wang, P., Gong, Y., Xie, H., & Liu, Y. (2016). SOMEDGRA: A case retrieval method for machine tool product configuration design. *Journal of Mechanical Science and Technology*, *30*(7), 3283-3293.

Wilke, W., & Bergmann, R. (1998). Techniques and knowledge used for adaptation during case-based problem solving. In *International Conference on Industrial, Engineering and*

Other Applications of Applied Intelligent Systems (pp. 497-506). Springer Berlin Heidelberg.

Wingate, G. (2016). *Computer systems validation: quality assurance, risk management, and regulatory compliance for pharmaceutical and healthcare companies*. CRC Press

Woodcock, J., 2004. The concept of pharmaceutical quality. American Pharmaceutical Review 7 (60), 10–15.

Xia, Q., & Rao, M. (1999). Dynamic case-based reasoning for process operation support systems. *Engineering Applications of Artificial Intelligence*, *12*(3), 343-361.

Xu, G., Zhang, Y., & Li, L. (2011). Web content mining. Web Mining and Social Networking, 71-87.

Yao, L., Chen, Z. C., & Yang, J. J. (2014). Research on risk assessment of ship repair based on case-based reasoning. *In Knowledge Engineering and Management* (pp. 53-59). Springer Berlin Heidelberg.

Yu, L. X. (2008). Pharmaceutical quality by design: product and process development, understanding, and control. *Pharmaceutical Research*, 25(4), 781-791.

Zhang, Y., Louvieris, P., & Petrou, M. (2007). Case-based reasoning adaptation for high dimensional solution space. *Case-Based Reasoning Research and Development*, 149-163.