Robotics and Computer-Integrated Manufacturing 29 (2013) 95–103

Contents lists available at SciVerse ScienceDirect

Robotics and Computer-Integrated Manufacturing

journal homepage: www.elsevier.com/locate/rcim

Continuous improvement of knowledge management systems using Six Sigma methodology

ChiaJou Lin [a,*], F. Frank Chen [b], Hung-da Wan [b], Yuh Min Chen [a], Glenn Kuriger [b]

[a] Institute of Manufacturing Information and Systems, National Cheng Kung University, 1 Ta-Sheuh Road, 701 Tainan, Taiwan
[b] Department of Mechanical Engineering, Center for Advanced Manufacturing and Lean Systems, University of Texas at San Antonio, One UTSA Circle, San Antonio, TX 78249-0670, USA

ARTICLE INFO

Article history:
Received 10 February 2012
Received in revised form
23 April 2012
Accepted 29 April 2012
Available online 11 August 2012

Keywords:
Six sigma
Knowledge retrieval evaluation

ABSTRACT

Knowledge retrieval is a decisive part of the performance of a knowledge management system. In order to enhance retrieval accuracy, an effective performance evaluation mechanism is necessary. Nowadays, there is not a standard evaluation framework for knowledge retrieval evaluation, because the evaluation set up is still technology-dependent, focusing on specific elements of the search context. The laboratory-based evaluation is not suitable to evaluate the knowledge retrieval process, since knowledge is dynamic, constantly changing and evolving. Besides, ambiguous query is also an important factor for the performance of knowledge retrieval systems. In order to improve the performance of knowledge retrieval, this paper proposes an evaluation mechanism using Six Sigma methodology to help developers continuously control the knowledge retrieval process. Specifically, this study involves the following tasks: (i) proposes a general knowledge retrieval framework based on the analysis result of knowledge retrieval, (ii) designs the knowledge retrieval evaluation framework using Six Sigma's Define-Measure-Analyze-Improve-Control (DMAIC) process and (iii) develops the related technologies to implement the knowledge retrieval evaluation mechanism. The knowledge retrieval evaluation mechanism allows system developers to maintain the knowledge retrieval system with ease and meanwhile enhance the accuracy.

1. Introduction

Knowledge has been recognized as the key resources of business survival and success in knowledge economy. Knowledge, while made up of data and information, can be thought of as much greater understanding of a situation, relationships, causal phenomena, and the theories and rules (both explicit and implicit) that underlie a given domain or problem [1]. Easier access to data and documents can help firms reduce the development cycles and lead times [2]. Therefore, knowledge management which consists of *Create*, *Storage*, *Retrieval*, *Transfer* and *Reuse* of knowledge has become an important approach to improve the competitive advantage of enterprises [3].

To transmit the right knowledge to the right people at the right time, knowledge retrieval is the major part of knowledge management. However, knowledge retrieval is a time-consuming task in the large knowledge base. It is impossible for users to filter all knowledge to determine the needed knowledge in large-scale knowledge bases, which involves a large amount of knowledge [4], and it is difficult for users to decide what knowledge is needed before they know it. Knowledge Retrieval Systems (KRS) which is the useful retrieval systems for supporting knowledge discovery, organization, storage and retrieval, guarantees access to large corpora of unstructured data [5]. In order to enhance retrieval accuracy, performance evaluation is also an important task for knowledge retrieval systems.

Nowadays, there is not a standard evaluation framework for knowledge retrieval evaluation, because the evaluation set up is still technology-dependent, focusing on specific elements of the search context [6]. Knowledge retrieval evaluation usually refers to the methods of information retrieval evaluation. The most commonly used methodology of information retrieval systems evaluation needs a test collection, which contains a document collection, a set of topical queries and a set of relevance assessments identifying the documents that are topically relevant to each query [6]. However, knowledge can be represented in different ways and stored in different types. Knowledge is dynamic, constantly changing and evolving [7]. Knowledge repository is under continuous growth in the real world. Hence, real search are often out of the evaluation scope. Besides, ambiguous query usually results in the lower accuracy of knowledge retrieval. Variance between queries is larger than the

* Corresponding author.
E-mail address: chiajou0101@gmail.com (C. Lin).

0736-5845/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.rcim.2012.04.018

variance between systems [8]. Therefore, user queries processing is also an important factor influencing the performance of knowledge retrieval.

Six Sigma, a customer-focused and data-driven quality strategy, is a rigorous and systematic methodology that utilizes collected information and statistical analysis to measure and improve performance. The philosophy of Six Sigma is to keep a process within its limits so almost no defects occur [9]. In order to improve the performance of knowledge retrieval, this paper proposes a knowledge retrieval evaluation mechanism using Six Sigma to help developers to continuously control the knowledge retrieval process. To achieve this objective, this paper first proposes a general knowledge retrieval framework based on the analysis result of knowledge retrieval and then designs the knowledge retrieval evaluation framework using Six Sigma's Define-Measure-Analyze-Improve-Control (DMAIC) process to continuously improve the efficiency of knowledge retrieval. Finally, this paper develops the related technologies to implement the knowledge retrieval evaluation mechanism. The evaluation mechanism is an effective tool to facilitate the knowledge retrieval process more robust and, hence, ensure satisfactory performance in support of problem solving, decision making, and knowledge innovation.

2. Related works

This section surveys numerous studies related to the aims of this paper, including knowledge retrieval, knowledge retrieval evaluation and Six Sigma.

2.1. Knowledge retrieval

Knowledge searching maps the constraints inputted by users with the knowledge repository. Rule reasoning [10] is the easiest way to search knowledge using one-on-one mapping constraints. However, different knowledge can satisfy one constraint while a piece of knowledge may satisfy multiple constraints. The mapping relationship between constraints and knowledge is implicit and unknown [11]. Therefore, the constraint driven knowledge searching method (CDKSM) using ANNs, the fuzzy theory and rule reasoning, is proposed to efficiently retrieve knowledge from Internet-based distributive knowledge integrated system (DKIS) repositories [11].

With the rapid growth of information and internet technologies, enormous volume of knowledge is created on the Web and in online databases. Most knowledge stored on the Web is un-structured. Hence, various engine technologies have been investigated. A keyword algorithm is a simple way to effectively present search results [12,13]. Jian et al. [14] proposed a knowledge search methodology using intelligent agents and information retrieval technologies to transform passive search and retrieval engines into active, personal assistants for the product development process. Jr et al. [15] presented an ontology-based model to represent context and rules for supporting automatic query expansion in a knowledge search in collaborative networked organizations (CNOs). Heeptaisong and Srivihok [16] created soil ontology to aid in soil knowledge search. User characteristics such as cognitive or thinking style [17–19], and search experiences and skill levels [20] also affect knowledge retrieval.

2.2. Knowledge retrieval evaluation

Information retrieval is the necessary technology for knowledge retrieval to guarantee access from large corpora of unstructured data. Traditional information retrieval evaluation is usually laboratory-based evaluation and relies on a test collection, evaluation measures and a comparative evaluation between algorithms, models or systems [6].

The information retrieval evaluation is a time-consuming and expensive process. A set of the relevance assessment in a test collection is needed to be judged by experts. However, the relevance judgments cannot be complete for any additional information because of the collection's growth. In order to achieve scalable assessment in the large collections, a pooling method [21] (i.e., the relevance is assessed over a subset of the collection formed from the top k documents returned by a number of different IR systems) is used to collect a set of documents and to reduce the manual effort. The pooling method provides a common basis for information retrieval to be reproducible and comparable [22].

The most important basic measures in information retrieval are recall and precision [8]. Recall is the proportion of documents retrieved by a search engine among all relevant ones, while precision is the proportion of relevant documents among all documents retrieved by the engine. Precision and recall are single-value metrics based on the whole list of documents returned by the system. In order to summarize performance, F-measure [23] and the mean average precision (MAP) [24] have been utilized to express the quality of a system in one number. However, those criteria are proposed for the non-ranked search results. Accordingly, Average Precision, R-Precision, Discounted Cumulative Gain and Weighted Mean Reciprocal Rank have been utilized to evaluation retrieval systems which present ranked results [25].

Besides, when databases are large, it is more difficult to evaluate the retrieval systems, because the relevance judgments become more and more difficult to complete. Therefore, Buckley and Voorhees [26] proposed a new measure, binary preference (bpref), and Aslam and Yilmaz [27] proposed three different evaluation measures, namely induced average precision (indAP), subcollection average precision (subAP), and inferred average precision (infAP), which are more robust to incomplete relevance judgments. Currently, several evaluation initiatives, such as TREC (Text Retrieval Conference), CLEF (Cross-Language Evaluation Forum) and NTCIR (NII Test Collection for Information Retrieval), are concerned with diverse retrieval applications, innovative usage scenarios and different aspects of system performance [8].

2.3. Six Sigma

Six Sigma has been developed by Motorola in the 1980s with focus on application in manufacturing environments. In the last decade, Six Sigma has received considerable attention in global companies to generate maximum business benefit and competitive advantage [28,29]. Six Sigma is defined as an organized and systematic method for strategic process improvement and new product and service development that relies on statistical and scientific methods to make dramatic reductions in customer-defined defect rates [30]. A Six Sigma process will approach 'zero defects', with only 3.4 defects per million opportunities (DPMO) for a defect to occur. Six Sigma differs from other quality programmes in its 'top-down' drive and its rigorous methodology that demands detailed analysis, fact-based decisions, and a control plan to ensure ongoing quality control of a process [31]. Besides, Six Sigma methodology enables practitioners to accurately remove hindering issues and demonstrate the improvements using statistical tools such as Pareto Chart and control charts [32,33].

Six Sigma uses a structured method, the Define-Measure-Analyze-Improve-Control (DMAIC) process, as the five steps for