Preliminary Overview of Data Mining Technology for Knowledge Management System in Institutions of Higher Learning

Muslihah Wook, Zawiyah M. Yusof, and Mohd Zakree Ahmad Nazri

Abstract—Data mining has been integrated into application systems to enhance the quality of the decision-making process. This study aims to focus on the integration of data mining technology and Knowledge Management System (KMS), due to the ability of data mining technology to create useful knowledge from large volumes of data. Meanwhile, KMS vitally support the creation and use of knowledge. The integration of data mining technology and KMS are popularly used in business for enhancing and sustaining organizational performance. However, there is a lack of studies that applied data mining technology and KMS in the education sector; particularly students’ academic performance since this could reflect the IHL performance. Realizing its importance, this study seeks to integrate data mining technology and KMS to promote an effective management of knowledge within IHLs. Several concepts from literature are adapted, for proposing the new integrative data mining technology and KMS framework to an IHL.

Keywords—Data mining, Institutions of Higher Learning, Knowledge Management System, Students’ academic performance.

I. INTRODUCTION

RECENTLY, Knowledge Management Systems (KMS) have been successfully applied in many areas (especially in business organizations) and also in education institutions, to achieve the significant target of a quality education process [1]. Reference [2] argues that Institutions of Higher Learning (IHL) are similar to knowledge businesses, in that both are involved in knowledge creation, dissemination, and learning. However, people in business world concerned with the profit they could gain by exploiting knowledge through the implementation of KMS whereas IHL consider that KMS could improve the quality of service deliveries and sustained competitive advantages in the academic world [3]. For that reason, KMS is seen as one of the most crucial management tool to transform knowledge resources in organization into intellectual capital for competitive advantage [4]. Thus, each organization must endeavour to align the Knowledge Management (KM) processes and organizational operations with appropriate technology that could create performance improvement environment [3].

Meanwhile, [5] stresses that KM effort is only a waste of time, if there is no technological infrastructure to support KMS. It will also fail if people are not willing to share their knowledge, no matter how easy it is to input it into a KMS. Reference [6] describe that technology plays an important role in KM, but it cannot be over emphasized. They point out that the core of KM is people whereas technology is just an auxiliary tool [7]. However, [8] stress that technology is still an essential tool for implementing KM processes and for support of all people activities in organization. In the light of this, [9] assert that KMS design should encompass organizational (processes), social (people and their behavior) and technological (infrastructure) aspects to effectively implement the practice of KM within organization. Thus, there is a need to successfully coordinate; people, processes, and technology, commonly known as pillars, using KMS [10].

Basically, people are the main actor for deciding the KM activities including the use of technology to operate the KM processes. Processes define the role of knowledge needed by people as well as determine the need for technology. Whereas, technology makes the KM processes possible and provides support for people involved in KM activities. Although, these three pillars are significant to accomplish a successful implementation of KMS within organization, this study focuses on technological aspect without considering that technology is the most important element compared to people and process. These authors seek to identify how technology can aid and enhance knowledge within organization by adopting data mining technology for implementing KMS in IHL context.

II. PROBLEM STATEMENT

Many technologies can be used to support the implementation of KMS [6], [11]-[15]. These technologies ranges from simple to complex ones [16] and could be classify into information technologies and web-based technologies [17]. Some of the information technologies are the databases, data warehousing, decision support systems, intelligent agents and data mining. Whereas, web-based technologies consist of internet, intranet, web 2.0, wikis, blogs, email, groupware and many more. Each of these technologies has different function and ability to support KM. However, studies on how those technologies could realize the KM processes in the context of IHL and how it could be integrated are at scarce [3].
Indeed, the successful development of KMS depends on the right selection of KM technologies. Data mining technology has been suggested as a promising technology for KMS implementation in IHLs [7], [18], [19] despite the increasing number of other sophisticated technologies. This is due to the ability of the technology to extract useful knowledge from large volumes of data from databases which is part of KM initiatives [20]. Moreover, data mining technology able to classify and predict future outcome based on previous data that could make an important contribution for KM processes; particularly in decision making process [21]. In this case, applied data mining technology to huge amount of data could give greatest advantages to IHL especially when IHL are no longer simply providing knowledge to students, but are also able to manage and expand their existing knowledge for future reference [22].

Even though data mining technology has been explored and evaluated in IHLs in various countries, there is only a handful of research on how to capture knowledge from the source and manage the knowledge produced. In fact, studies on data mining have emphasized more on testing the ability of the algorithm to produce an accurate model, rather than on developing a comprehensive application system for education users to make decisions [23]. In other words, there is a lack of studies on the integration of data mining technology in the context of KMS for IHLs. Therefore, this study seeks to identify the potential use of data mining technologies in KMS; particularly for monitoring students’ academic performance in IHLs.

III. INSTITUTIONS OF HIGHER LEARNING

Each IHL is unique in its objective, scope, size, and priority [24]. However, all IHLs basically have the same mission, primarily “to train capable manpower to formulate, plan, manage, and run the country’s development programs, and to carry out research to solve the various problems that confront society” [25]. To achieve this mission, IHLs must also be able to rely upon internal core competencies and external pressures [26]. The internal challenges of an IHL is definitely the management processes of students, staff, academic, and research that indirectly produce knowledge during their activities; whereas the external pressures are such as the influence of government, other IHLs, ranking and technological aspects that can impact greatly on competitive advantage. Fig. 1 summarizes the challenges and pressures faced by IHLs.

Fig. 1 The Challenges and Pressures in IHLs

Generally, students are the internal core of any IHL. Even though, the staff (including teachers, advisors, lecturers, and administrators), academic services and learning as well as research are also important to IHLs, but, this study focuses only on students since they are the product of IHLs. Hence, the IHL’s staff should focus on the management of students by developing strategies for the generation and utilization of their data, information and knowledge towards enhancing and sustaining the organizations’ performance [26]. An IHL’s performance is undeniably determined by the number of students enrolling and graduating, as well as its innovative researches. As students are the main resource of IHLs, their academic performance was the most interest to researchers from 1993 to 1999 [23]. In fact, based on the linear regression models, which were used by [27] to find out which variables are more likely to be associated with students, it was discovered that academic performance is the most important variable. Therefore, information on the students’ academic performance is seen as important knowledge that must be explored and managed in a systematic way, to understand the needs and preferences of students during academic courses in IHLs.

Meanwhile, although all the external pressures from government, other IHLs and ranking are vital to IHLs, these pressures could be overcome with the aid of technologies for enhancing the performance. Furthermore, the advanced technologies which rapidly changing in external environment are the most influential to IHL; particularly in the era of Information and Communication Technologies (ICT) [25], [26]. Some of the trends in this environment include the growth of data and information over the web, databases, data warehouses, and repositories. For example, each year, IHLs produce internally generated data and information from the organization’s operations, such as student records, courses, faculty, and staff. Although most IHLs have adopted tremendous and superior database within each department or unit, they actually had “islands” of unconnected databases [11]. For instance, many IHLs in Malaysia are not fullest utilizing knowledge to improve their performance; this is because the data, information and knowledge available in the IHL are not properly managed [28]. Additionally, [2] argues that many IHLs have created sophisticated library and information systems. However, their implementation is often more information driven than knowledge driven. Therefore, there is an urgent need to change these current practices of information driven, into knowledge driven, which is seem able to meet the urgent need of cultivating high-level, modern, and comprehensive intellectuals in this ‘net’ age [29].

Knowledge is viewed as a strategic input that is vital to any educational institution, for improving the quality of the education process [2], [30]. In the education field, knowledge has become the most important factor to determine the quality of education as well as to develop quality graduates to reach the nation’s goal of building a knowledge-based society (K-society) [7], [28]. For this reason, knowledge within an IHL needs to be identified, transformed, and disseminated effectively [31]. To do this, there needs to be a tool, a
technology, a mechanism, or an enabler, that is able to create, manage, and use existing knowledge stores better, for preparing the organization to meet new accountability, effectiveness, and efficiency requirements [32]. According to [3] and [33], there is an urgent need to link KMS with IHL processes to improve outcomes in this contemporary knowledge economic time. Therefore, KMS are seen as the best choice for helping IHLs to improve their educational service quality [34].

IV. KNOWLEDGE MANAGEMENT SYSTEMS

KMS are a class of information system that are IT-based, and developed to support creation, storage/retrieval, transfer, and the application of knowledge in organizations [13]. Meanwhile, other researchers have defined KMS as a tool in realization of KM for helping an organization in problem solving and decision-making activities [35], [36]. The advancement of ICT can be used to enhance, systematize, and expend the valuable knowledge intra and inter organizations. However, there is no single ICT that can facilitate for the whole processes of KM in organization [21]. In fact, [15] stress that every KM process needs corresponding technological support since technology relevant to KM does not exist individually. There is a need of dynamic technology, which will continuously change simultaneously with knowledge theory and application [15]. Therefore, this study focuses on the technologies that are able to support towards selected processes in the management of knowledge within an organization, without emphasising that it is the most important element of KM. These authors prefer to emphasize that technology as the enabler for implementing a system for managing the processes of knowledge in an organization.

Before attempting to implement a KMS, an organization needs to undertake the strategies or blueprints of the KM process. Some studies have proposed different KM processes, such as: [37] who propose the six different processes of knowledge creation, capture, organization, storage, dissemination, and application; [38] focuses on knowledge acquisition, creation, utilization, and sharing; [39] suggest knowledge creation, acquisition, organization, retention, dissemination, and utilization; [14] focus on knowledge generation, storing, codification, representation, transformation, use, transfer, sharing, retrieval, access, and searching; [40] introduces eight phases of the KM lifecycle, which include knowledge creation or acquisition, knowledge modification, use, archiving, transfer, translating/repurposing, user access, and disposition; and [41] classify KM processes into creation, dissemination, and incorporation.

It is important to understand that KM is a complex process, which will be understood and used differently by people in different contexts [2], because they like to cover all activities of managing different knowledge types, within distinct organizational types, but conveying similar meanings [39]. For example, some people may think that the core processes of KM are knowledge creation, generation, sharing, use, and dissemination. But others consider knowledge acquisition, codification, generation, and representation to be more important. In other words, sometimes all of the terms that represent the same thing are duplicated and overlap. For example, knowledge generation and creation, knowledge acquisition and retrieval and knowledge use and utilization.

Although there are many KM processes, but only three is considered in this paper. As such, this study attempts to propose a framework that consist only three processes as shown in Table I. These processes are modified from [38] and [40]. Table I also shows that there are many types of technologies that can support knowledge creation, utilization, and disposition processes (modified from [40]). As mentioned earlier, this study attempts to integrate data mining technology with KMS, to monitor students’ academic performance in IHL context since [7], [18], [19] suggest that this technology is promising. Meanwhile, a web technology and a decision support technology are used for the knowledge utilization process, and a database technology is used for the knowledge disposition process. All of these technologies are to be integrated to form a KMS to monitor students’ academic performance in IHLs. This view is consistent with [42] which more integrated systems are needed that support not only single process but also other processes in the KM cycle.

<table>
<thead>
<tr>
<th>KM Processes</th>
<th>Enabling Technologies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge creation</td>
<td>Authoring tools, interface tools, data capture tools, decision support tools, simulations, professional databases, application-specific programs, database tools, data mining tools, pattern matching, groupware, controlled vocabularies, storage media, graphics tools</td>
</tr>
<tr>
<td>Knowledge utilization</td>
<td>Interface tools, visualization tools, decision support tools, simulations, application-specific programs, database tools, pattern matching, groupware, infrastructure, web tools</td>
</tr>
<tr>
<td>Knowledge disposition</td>
<td>Database tools, storage media</td>
</tr>
</tbody>
</table>

V. DATA MINING IN IHLs

Reference [43] assert that knowledge could be generated by using data mining technology. Data mining refers to a process and set of techniques that can be used to extract useful information and knowledge from large volumes of data, which are stored in databases, data warehouses, or as other explicit knowledge [44], [45]. It is also a particular step in the Knowledge Discovery of Databases (KDD) process that aims to extract unknown (but potentially useful) information and knowledge from incomplete, ambiguous, and random data [45], [46]. According to [47], the conversion of knowledge from implicit to explicit takes place as a result of the KDD process, due to the fact that all knowledge in a knowledge base is already explicit, and the KDD process creates new knowledge based on it. It is also claimed that the conversion of knowledge from explicit to tacit occurs when people interpret new knowledge from different perspectives.

Recently, there has been an increasing research interest in data mining in IHLs. Studies by [48] and [49] reveal that this kind of research concerns knowledge extraction from educational data known as Educational Data Mining (EDM).
Educational data is in the form of historical and operational data that has been collected and recorded in an institution’s records or databases. Universities collect large amounts of student data for years. For example: enrolments, curriculum and co-curriculum performance, graduation, alumni, career path, and others. Thus, current universities are rich in data, but poor in information [50]. As large amounts of data are stored within these databases, conventional statistical analysis and database management tools are no longer adequate to analyze this large amount of data [51]. In fact, the administrations are always make mistakes particularly in making decision on students’ performance and behaviours toward education.

To overcome this problem, a proper method of extracting knowledge from IHLs databases is required to improve their decision making process. Several studies have applied various kinds of data mining techniques in IHLs such as decision tree, neural network, naïve Bayesian, k-nearest and many others [49]. Based on Table II, the main operation in IHL is academic performance and most of the techniques used to analyze students’ academic performance is the best model from the previously tested algorithm. For example; neural network [52], decision tree [44], [49], [53] and combining more than one technique; known as a hybrid technique [54]-[56]. Based on these studies, the knowledge created by data mining techniques enable IHLs to make better decisions, advanced planning in directing students and predicting students behaviour with higher accuracy [57].

### TABLE II

<table>
<thead>
<tr>
<th>Data Mining Techniques</th>
<th>Researchers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neural Networks</td>
<td>[52]</td>
</tr>
<tr>
<td>Decision Tree</td>
<td>[49], [53], [58], [59], [60], [61], [62]</td>
</tr>
<tr>
<td>Rough Set Theory</td>
<td>[63], [64], [65], [66], [67]</td>
</tr>
<tr>
<td>Clustering</td>
<td>[68], [69]</td>
</tr>
<tr>
<td>Association rule</td>
<td>[50], [54], [55], [56], [70], [71]</td>
</tr>
<tr>
<td>Hybrid</td>
<td></td>
</tr>
</tbody>
</table>

Even though each data mining techniques as shown in Table II is a robust technique, however, but users without data mining background face problem to understand the technique. In fact, the application of educational data mining has to be designed in an easier form; especially for educators and non-expert users of data mining [23]. According to [23], “most data mining tools are too complex for educators to use, and their features go well beyond the scope of what an educator may want to do.” Therefore, there is a need for designing a meaningful application system to facilitate IHLs users to build their valuable knowledge, which would lead to a better knowledge-driven decision [53].

### VI. FRAMEWORK

Various frameworks have been proposed to serve as blueprints of how to represent a KMS. Fig. 2 shows the proposed framework of this study, comprises of knowledge creation, utilization, and disposition modules. According to [13], data mining is a technology that supports knowledge creation process in a KMS. They also identifies that this technology enable the merging, categorizing, reclassifying and synthesizing of explicit knowledge into new explicit knowledge. Thus, the knowledge creation module involves the knowledge discovery process, which consists of data selection, pre-processing, transformation, data mining algorithm, and model analysis. The output (or knowledge) from this module depends on the types of data and data mining techniques used. As this study focuses on students’ academic performance in IHLs, the data was collected from the students’ database. The attributes selection depends on academics (i.e., grades in every semester and courses taken) and student’s background (i.e., demographics, education, and personality). The students’ data (according to its attribute) will then be mined using the best techniques from previously tested algorithms to uncover valuable knowledge that can be utilized by IHL’s staff for further action.

![Fig. 2 Proposed KMS Framework Using Data Mining Technology to Monitor Students’ Academic Performance in IHL](image-url)
consists of a sharing process and a decision support process. Sharing process refers to the distribution of knowledge and people are encouraged to share and reuse knowledge within the organization [36]. Knowledge utilization and sharing should take place simultaneously, to generate added-value amongst the IHLs community of practice, such as administration staff, advisors and lecturers; in the form of tacit and explicit knowledge. This activity can be done by adopting a web tool that is able to disseminate knowledge to the right person at the right time. However, the decision support technology is the enabling technology, for the decision support process to assist and resolve problems. For example, problems of high numbers of students failing at certain courses, determining specialization, scholastic abandonment, and many more. Therefore, both sharing and decision support processes depend on the data mining technology’s knowledge creation module to realize the maximum utilization of knowledge in IHLs.

The knowledge disposition module is the final destination of knowledge created during the earlier stages. In this module, the explicit knowledge that was stored in the database technology is be either destructed (for valueless knowledge) or preserved (for valuable knowledge). Even though all knowledge created in the data mining process may be valuable to someone at some point in the future, from a practical perspective, knowledge with limited future value is discarded to save space in the database and reduce overheads [40]. For example, knowledge on students’ academic performance for previous 10 years can be either destructed if no longer important or preserved depends on the value it carries. Therefore, the method of how to destruct or preserve explicit knowledge should conform to the IHL’s policy.

Based on the explanations as the above, the primary goal of the framework is to provide IHLs with valuable knowledge that can enhances all phases of knowledge. In order to accomplish this goal, the data mining technology is proposed to extract useful information from large volumes of students' database. Data mining technology is believed to be able to find hidden knowledge from a massive of data which is impossible to be done manually. Additionally, the knowledge created should be managed in further phases (i.e. knowledge utilization and knowledge disposition) to see the continuation of knowledge within organization.

VII. CONCLUSION

It is acknowledged that data mining technology gains its importance and accepted as the most promising technology for KMS; particularly in the context of IHLs. Data mining technology able to create valuable knowledge to be used in helping decision making to eliminate repeating previous mistakes of analyzing students’ data. Thus, the integration of data mining technology and KMS could also help to enhance accountability, transparency and the smooth running of IHL.

REFERENCES


