Knowledge Discovery in Database: A Knowledge Management Strategic Approach

AkanmuSemiu Ayobami[1] **and Shamsuddeen Rabi'u**[2]

[1]*Universiti Utara Malaysia, Malaysia, ayobami.sm@gmail.com*
[2]*Federal College of Education, Katsina, Nigeria, shamsujaja@yahoo.com*

ABSTRACT

Knowledge management had been a critical focus and interest in Information Technology, especially as it affects business organizations through the implementation of business intelligence and expertise. Knowledge discovery and knowledge conversion (tacit/implicit to explicit knowledge) play important roles in these aspects; through the application of technologies in the SECI model to aid knowledge management, and identifying the sources of the expertise whether in humans or physical databases serve as the basis for expertise's knowledge management. This paper presents in detail the significances of knowledge discovery in databases (KDD) in achieving an all encompassing knowledge management strategy. This strategy must comprise of transparent and multiple interrelationships of organizational agents through shared mental maps, collaborative and distributed technologies, and solves all problem in other ways with a special focus on data mining which is also found in the KDD process. Extensive literatures were reviewed tooperationalize Knowledge discovery in human and in data ware houses as its affect knowledge management, and bring to the fore the processes involved in KDD process, its applications, understanding using SECI model, possible challenges, and suggestthe future research areas to solve the observed challenges.

Keywords: Knowledge Management; Knowledge discovery in databases; SECI model.

I   INTRODUCTION

Knowledge is said to be the most unquantifiable asset to organisations, companies, and individuals (Ward and Peppard, 2002). The growth of any company, educational institutes and firms relies on the quality of the available knowledge in such companies, therefore, organisations must efficiently create, locate and share the knowledge (tacit and explicit) of the organisation and preserve its expertise in order to remain in the competitive business world (Zhan, 2008; Reinmoeller and Chong, 2002). Knowledge Management is a relative organisationalaction: Young (2010) gave the five-key steps that characterized the APO's Knowledge Management process as: identifying the knowledge, creating the knowledge, storing the knowledge, sharing the knowledge, and applying the knowledge. The identification and creation are undoubtedly synonymous to Knowledge discovery, and these form the basis of any productive knowledge management process in any organisation (Rohana et al., 2010).Zhao (2001) further highlighted, that Knowledge management is a study concerned with the way of exploring and developing the knowledge assets of an organisationwith an aim to further strengthen the organization's objectives.

Most importantly in knowledge discovery processes is the creation of knowledge systems which could be supported by the use of data mining technologies, especially for the uncovering of new relationships among the explicit data; serving as model which will be used to predict, categorize and implement information as a highly valuable assets in business intelligence (Yi et al., 2006). This invariably led to the overlapping meanings, hence a need for distinct understanding of data mining and knowledge discovery in databases (KDD).

In the light of this: Section 2 of this paper will beam its light on Knowledge Discovery in Databases via Knowledge Capturing, Knowledge Creation with reference to SECI model, and lastly Knowledge Creation Framework. Section 3 will x-ray the similarities and differences between KDD and Data mining, and lastly their applications in the Knowledge Management processes and technologies; while section 4 discusses the emerging issues and present trend of research work that are applicable to KDD as a Knowledge Management strategy, and the conclusion is precisely but succinctly given in Section 5.

II   KNOWLEDGE DISCOVERY IN DATABASES (KDD)

Knowledge discovery in Databases is an intersection of several disciplines like statistics, databases, AI, visualization, and equally a link to high performance of parallel computing; and it affects an interdisciplinary knowledge domain and other general-

purpose tools (Seifert, 2004).Luo(2008)explained that Knowledge Discovery in Database (KDD) consists of the following processes: data selection, data cleaning, data transformation, pattern searching which are also called data mining, finding presentation, findingsand evaluation. Figure I. illustrates the processes in Knowledge Discovery in Databases.
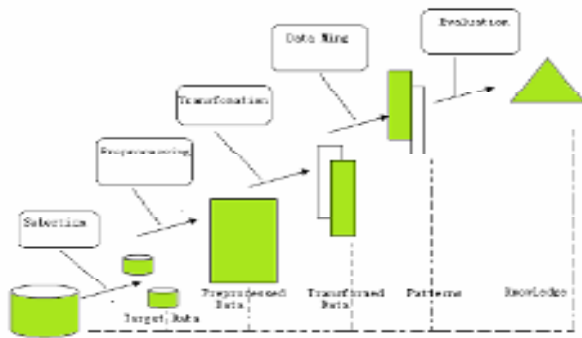

Figure I: Processes in Knowledge Discovery in Databases. (Adapted from Lou, 2008)

The above-given explanatory piece gives an exclusive view, positing databases solely from the data warehouses perception.

A   Knowledge Discovery

Eugene et al. (2010) defines Knowledge discovery as the nontrivial extraction of implicit of an information that is previously unknown but of potential usefulness extracting the information from the available data. These knowledge discovery systems rely on technological hardware and technologies which support both the socialization and combination processes without any need to distinguish the creation and discovery phases of the knowledge, but only considers their similarity in terms of meaning; especially as innovation and as advancement of knowledge (Seifert, 2004).

Rohanaet al. (2010) gave an example of sessions of organisations' brainstorming as a good avenue to detect and discover new knowledge which may not be felt individually but through collective activity. In these instances,the multiple embodiments of tacit/explicit knowledge are integrated to achieve a new but more complex type of the knowledge. In achieving this, there is need to re-structure the existing explicit knowledge so as to produce a new brand type.Knowledge discovery mechanisms and technologies are used in the facilitation of socialization and combination either within an organisation or across organisations; situations

whereby expertise in the organisations are analogical to the data warehouses.

B   Knowledge Creation (SECI Model)

In knowledge creation, Socialization is defined as the process of bringing tacit knowledge extracted from many individuals to a joint point of utilization through collective activity instead of being through a written or verbal instruction, knowledge discovery of tacit form is then enabled by activities between masters and apprentices, among researchers in conferences as the case may be (Schreiber, 2001).
SECI Model explains sequentially, cyclically, systematically and in details the process of creating this knowledge. Allan (2010) citing Tsoukas&Valdimirou acknowledged the organisational sentiment echoed by the shift of explicit to tacit knowledge, but the manifestation of SECI model is a product of a four-phased methodological approach to knowledge creation and knowledge transfer theory. Figure II illustrates the SECI model diagram.
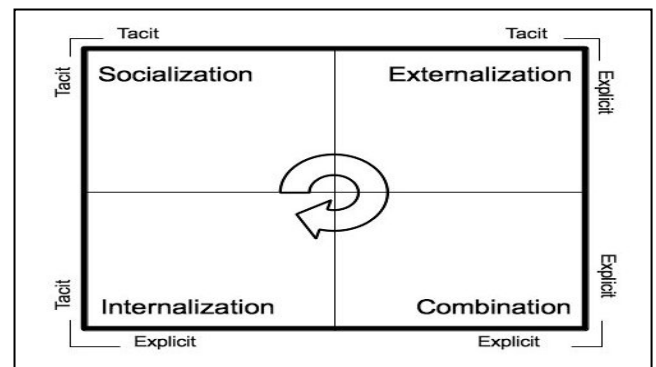

Figure II: Conceptual diagram of the SECI Model

Most importantly in knowledge discovery is that the knowledge creation systems could be supported by the use of data mining technologies, especially for the uncovering of new relationships among the explicit data; serving as model which will be used to predict, categorize and implement information as a highly valuable assets in business intelligence (Yi et al., 2006).
However, according to the SECI model, knowledge is continuously converted and created as users use the discovered knowledge by practice and learning; a model that is in a cyclical order and must be well taken into account for a result-driven and effective knowledge management.Discussing these phases sequentially and briefly:

Socialization
This consists of peculiar ways of sharing tacit knowledge with others by the means of mentoring,

discussion and apprenticeship. It is simply explained by employing a knowledge sharing process of socialization: mentoring, imitation, observation and through practice, all with the primary motive of resulting in sharing of knowledge (Armit,2001).At this present technological age, the need for the adoption of technologies to aid the process of tacit knowledge sharing through socialization is unparalleled; with the virtual environments. The social networking sites are virtual communities that exist online; allowing different groups of people to come together in line with their shared interests, be itpolitical activism or causes with the aim of sharing these ideas, among other reasons(Hemetsberger and Reinhardt, 2004; Ellison et al., 2009). Asleema (2010) agreed that it is done for day to day sharing of knowledge between members; providing access to key individuals.

The common purpose of all the above-stated technologies in the Socialization phase of the SECI model is that the technologies are employed in order to build a collaborative and sharing platform of the tacit knowledge (Fadhilahet al.,2010).

Externalization

This is the process through which tacit knowledge is being converted to explicit knowledge. The knowledge conversion is said to be externally done when dialogue transforms tacit knowledge into explicit knowledge. This also consists of many effective and useful technologies; among them are Weblog, Instant messaging, Wiki and VoIP (Becerra-Fernandez, 2004).

Becerra-Fernandez (2004) further defined Weblog or Blog in its shortened form as a web-based publication which primarily consists of periodic articles, and Wikis as other types of websites where users are given rights to create, edit and delete content.
A well known potential application of Wiki is its open knowledge exchange systems which is shown in the evolution of Wikipedia; a multilingual project tha creates a complete and accurate content c encyclopedia in the open source communit (Hemetsberger and Reinhardt, 2004). They are used t develop and update information that is useful for othe many users who only hold parts of the combine knowledge body individually.

Combination

In Combination, the knowledge conversio isdone by combining different types of explic knowledge once the knowledge is captured(Rice an Rice, 2002), this is also done through the use of technologies.RSS is an outstanding technology in this

respect; making resource sharing easy across networks, it directs the right knowledge to the corresponding right people bringing contents from different sources to a learner's personal space.The content of the message is made valuable by the responsibility of collective intelligence through filtering, rating, feedback, reviews and supports of the certification of the people's expertise and the evaluation of individual digital reputation (Liu et al.). Examples of this collective intelligence are Amazon's review and recommendation system, YouTube's rating system, Google's Page-Rank algorithm and eBay's feedback.

Internalization

In Internalization, explicit knowledge is converted into tacit knowledge, consisting of 'learning by doing' process. This process of internalization occurs when the previous modes of knowledge conversion; socialization, externalization and combination are internalized in the minds of the people as tacit knowledge (Becerra-Fernandez, 2004;Tse et al., 2010).

This process equally brings learners together with the purpose of competitive and cooperative and collaborative relationship through multi-player and multi-user simulations which offer a potential to learn through a new form of social experience. (Tseet al.,2010).

Conclusively, Garnia (2002) posited that the core behavioural assumption in the SECI model is a continuous flow of knowledge between individuals, groups and organisations, thus creating a knowledge value through the synergies among the knowledge holders within the supportive and developmental organisational context. Figure III illustrates the key elements in the SECI model; I= Individual, G= Group, and O= Organisation.
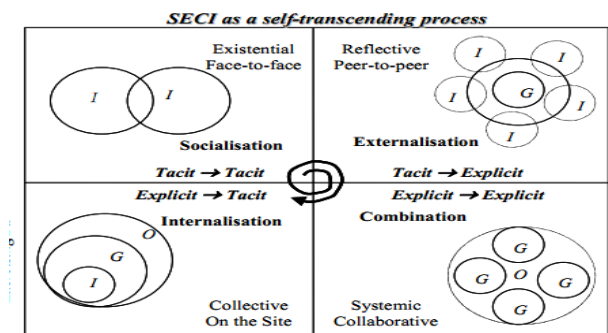

Figure III: SECI model's key elements (Adapted from Rice and Rice, 2002).

C   Knowledge Creation Framework

Notwithstanding, having explained the technologies involved in the Knowledge creation process, it is essential to equally understand the criteria that rule this knowledge-based competition so as to help in maintaining edge in the competitive knowledge creation environment. In this light of this, a knowledge creation framework is required:

Garnia (2002) explained that this framework must explain clearly the nature of the knowledge, the place where it is created and applied, and the mechanisms which allows the transfer of the knowledge. Three main issues were identified to be pursued in the creation of knowledge; these are:

1. The nature and typology of the knowledge based on the relevant and generally accepted criteria;

2. The determination of the different entities, levels, systems or agents which have the ability to create knowledge; and

3. The manner through which these entities can develop within themselves, capture and relate knowledge related to lower or higher levels.

Garnia(2002) posited that the classification of the knowledge must be done by a generally acceptable criterion. A carefully used and clear cut distinction between tacit and explicit knowledge proves that the tacit knowledge can establish a two-extreme ends; with one extreme showing the opposite of the other.

The academics are still on intensive research to position this appropriately, looking at the formation and conversion of the knowledge especially in knowing that if groups and teams are constituted by individuals that own knowledge-creation and learning abilities.

In summary, four basic levels are concluded to be the prerequisites for any sustainable knowledge creation to happen; they are: the individual level, the group level, the organisational level, and the other inter-organisational environmental level. The individual level is classified as the basic unit or element for knowledge creation, while the group, organisational and inter-organisational environmental levels are the higher level knowledge for creating systems (Garnia, 2002).

III KNOWLEDGE DISCOVERY AND DATA MINING

The possible use of data mining technologies for Knowledge discovery in databases often gives an overlapping understanding, though with similarities and differences (Yiet al., 2006).

A Application of Knowledge Discovery and Data Mining in Knowledge Management

Knowledge discovery in database (KDD) is the process used to search for and extract meaningful information from volumes of documents and data (Fayyad et al., 1996). These include tasks like knowledge extraction, data archaeology, data exploration, data pattern processing, data dredging, and information harvesting. Knowledge discovery in database involves finding, establishing and interpreting patterns from data, involving the application of different algorithms to interpret the patterns generated by these algorithms(Fayyad et al., 1996; Mudhol and Gowda, 2004).The common scenario in organizations is that data are often buried deep within very large corporate databases, data warehouses, text documents, or knowledge repositories, all of which may contain data and information. Knowledge discovery in database is also widely known as data mining (DM). Also, KDD is defined to involve all the phases of knowledge discovery including the application of DM techniques (Fayyad et al., 1996).

Data mining search for previously unknown information or relationships in large databases is one of the useful techniques for eliciting knowledge from databases, documents, e-mail, and so on(Fayyad et al., 1996). This process invariably preserve, hence, manage the knowledge domain of the organisation.

B   Data Mining Tasks

Due to the fact that many patterns are widely used in a large database,Garnia et al. (2002) highlighted the tasks of data mining as distinct and diverse, and can be classified accordingly as done below:

    a. Summarization: is the abstraction or generalization of data. A set of task-relevant data is summarized and abstracted. This results in a smaller set which gives a general overview of the data.

    b. Classification: This is constructed by analyzing the relationship between the attributes and the classes of the objects in the training set and can be used to classify future objects which help in developing a better understanding of the classes of the objects in the database.

    c. Clustering: It identifies classes also called clusters or groups for a set of objects whose classes are unknown. The objects are so clustered that the interclass similarities are

maximized and the interclass similarities are minimized.

d. Trend analysis: Time series data are records accumulated over time. For example, a company's sales, a customer's credit card transactions and stock prices are all time series data. Such data can be viewed as objects with an attribute time.

Data mining systems have made significant contributions, for example, in breast cancer diagnosis, electronic commerce (e-commerce) applications, financial planning and others. It helps to provide hard data ready for analysis, and provide organisations with an excellent opportunity to make profitable by using these techniques (Fayyad et al., 1996; Mudhol and Gowda, 2004).

The following are some of the example applications of DM to KM for business:

a. In marketing, predictive DM techniques, such as artificial neural networks are used for target marketing including market segmentation. This allows the marketing departments to segment customers according to basic demographic characteristics.

b. Retail: DM methods have been used for sales forecasting by taking into consideration multiple market variables, such as customer profiling based on their purchasing habits.

c. Insurance. DM techniques have been used for segmenting customer groups to determine premium pricing and to predict claim frequencies. Clustering techniques have also been applied to detecting claim fraud and to aid in customer retention.

a. Business applications: Many organizations now employ data mining as a secret weapon to keep or gain a competitive edge. Data mining has been used in database marketing, retail data analysis, stock selection, credit approval, etc.

b. Science applications: Data mining techniques have been used in astronomy, molecular biology, medicine, geology and many more.

c. Other applications: Data mining techniques have also been used in health care management, tax fraud detection, and money laundering monitoring and even sports (Shaw et al.2001; Fayyad et al., 1996).

IV ISSUES AND SUGGESTED RESEARCH AREAS

The implementation of Knowledge discovery as its affects Knowledge management in the processes of Knowledge capturing, identification and capturing is different from the view of knowledge discovery in data sets of data warehouses and repository, hence, positioning the clear-cut distinction becomes challenging.

Since tacit knowledge found in human could not be estimated based on mere certifications, the process of precisely discovering who among the organisational human ware has the needed knowledge which could aid the organizations' growth requires a further in depth study (Alavi and Leinder, 2009). Ulrike & Dorothy (2002) suggested a need to examine the importance of knowledge discovery and knowledge management in the organisation from the both the negative and positive perception, since it is asserted that knowledge could be seen as been a double-edged sword. Also, the growing trends of needed technological automation in business processes as its affects organisational context, knowledge taxonomy, ontology and so on areareasof urgent research focus (Nilakanta et al., 2006).Fabbriet al. (2011) observed that the need to apply data mining techniques in the development of information technologies for use in fields like health, security, business analysis, and forensics and so on is a great insight for researchers in this field.

V CONCLUSION

Knowledge being an unquantifiable asset to organisation needs preservation, retention and adequate knowledge management to avoid brain drain in the organisation, and provide a competitive advantage in the emerging and dynamic business world (Nilakanta et al., 2006). Following a logical suit, Knowledge discovery automatically serves as the starting point for any KM process either as its affects human (individuals, groups or organisations) or data warehouses which is also called data repository.

Considering the sources of the knowledge to be discovered and subjected to management process, SECI model is shown as a comprehensive illustration using these KM strategies with the aid of respective technologies for all thefour phases(Echmann, 2003), this is a direction pointed by this study: that knowledge discovery forms the basis and an important phase of knowledge management no matter the perception. Also of importance is the conventional understanding of knowledge Discovery in Database which is synonymous to Data mining. Its application

areas in business, science, medicine, and the academics are brought to the bear to underscore its significances and suggest its further exploration to solve the pending challenges and aid its full incorporation in order to achieve a more robust business-oriented structure (Echmann, 2003).

In conclusion, the challenging issues in the field of Organisational knowledge management, organisational memory, applicability of data mining in various fields and others (Alavi and Leinder, 2009;Ulrike and Dorothy, 2002) are examples of the present research areas which must be promptly attended to.

REFERENCES

Alavi, M. &Leidner, D. (2009). Knowledge Management and Knowledge Management Systems: Conceptual Foundation and Research Issues. France. INSEAD.

Allan, F. (2010).The SECI model. Available at: http://www.knowledge-management-tools.net/knowledge-conversion.html

Armit, T. (2001).The essential guide to Knowledge Management. New Jersey. Prentice Hall PTR. ISBN: 0-13-032000-5

Asleema, H. (2010). The Articulation of KM Strategy in Public Sector Organisations. Proceeding of Knowledge Management 5th International Conference, May, 2010. UUMCAS, Malaysia.

Becerra-Fernandez, I., Gonzalez, A., &Sabherwal, R. (2004).Knowledge Management – Challenges, Solutions, and Technologies. New Jersey: Prentice Hall. Retrieved from: http://www.worldscibooks.com/etextbook/5639/5639_chap01.pdf

Echmann, C. (2003) Investigation of technologies in the knowledge management context, School of Computing, Dublin Institute of Technology, Research Paper (ITSM), DIT, 2003.

Ellison, N.B., Lampe, C. &Steinfield, C. (2009). Social Network Sites and society: Current Trends and Future Possibilities. The Potential for Technology-Enabled Connections. Retrieved from: https://www.msu.edu/~steinfie/EllisonLampeSteinfield2009.pdf

Eugene, O., Khalil, M., Seyyeddali, Z., &Huam H. (2010).Effect of Individual Factors on Knowledge Sharing. Proceeding of Knowledge Management 5th International Conference, May, 2010. UUMCAS, Malaysia.

Fabbri, D., LeFevre, K., Hanauer, D.A. (2011). Explaining Accesses to Electronic Health Records.Proceedings of 17th ACM International Conference on Knowledge Discovery and Data Mining, USA.http://dx.doi.org/10.1145/2023582.2023585

Fadhilah, A., Ashley, A. (2010). Information Technology for Knowledge Management in Malaysia: the Effects of Organisational Factors. Proceeding of Knowledge Management 5th International Conference, May, 2010. UUMCAS, Malaysia.

Fayyad, U., Shapiro, G. & Smyth, P. (1996).From Data Mining to Knowledge Discovery in Databases.*American Association for Artificial Intelligence.*Retrieved from: http://www.kdnuggets.com/gpspubs/aimag-kdd-overview-1996-Fayyad.pdf

Garnia, M., Fernando, E., Martin, C., & Gregorio. (2002). The Knowledge-Creation Process: A critical examination of the SECI model. Available at :http://apollon1.alba.edu.gr/OKLC2002/Proceedings/pdf_files/ID151.pdf

Liu, H., Ramasubramanian, V. &Srer, E.G. Department of Computer Science, Cornell University, Ithaca. Retrieved from: http://www.cs.cornell.edu/people/egs/papers/rsssurvey.pdf

Lou, Q. (2008). Advancing Knowledge Discovery and Data Mining.School of Electrical and Information enginerring, WITN, China.*IEEE Computer Society.* 0-7695-3090-7/08

Mudhol, M. &Gowda, P. (2004).Data Mining in the Process of Knowledge Discovery in Digital Libraries.Inflibnet.Retrieved from:

http://ir.inflibnet.ac.in/dxml/bitstream/handle/1944/449/04Planner_21.pdf?sequence=1

Nilakanta, S.; Miller, L. & Zhu, D. (2006). Organizational Memory Management: Technological and Research Issues. *Journal of Database Management.* 17(3), 85-94. Retrieved from: http://www.igi-global.com/viewtitlesample.aspx?id=3349

Reinmoeller, P. & Chong, L. (2002).Managing the Knowledge-Creating Context: A Strategic Time Approach. USA. Blackwell Publishers.11 (3). Retrieved from: http://www.blackwellpublishing.com/content/BPL_Images/Journal_Samples/CAIM0963-1690~11~3~249%5C249.pdf

Rice, J.L. & Rice, B.S. (2002). The Applicability of SECI model to multi-organisational endeavours: An integrative review. *International Journal of Organisational Behaviour.*9(8), 671-682 ISSN 1440-5377. Retrieved from: http://www.usq.edu.au/extrafiles/business/journals/HRMJournal/InternationalArticles/Volume%209%20Knowledge%20Mgt/Vol9No8Rice.pdf

Rohana, N., Abdul Razak, I. (2010).The Effect of Knowledge Sharing on Organisational Performance in Small and Medium Enterprises. Proceeding of Knowledge Management 5th International Conference, May, 2010. UUMCAS, Malaysia.

Schreiber, G., Akkermans, H., Anjewierden, A., de Hoog, R., Shadbolt, N., de Velde, W.V., &Wielinga, B. (2001).*Knowledge Engineering and Management: The CommonKADS Methodology.* MIT Press: Cambridge.

Seifert, J. (2004). Data Mining: An Overview. Congressional Research Service.The Library of Congress. Retrieved from: http://www.fas.org/irp/crs/RL31798.pdf

Shaw, M.J., Subramaniam, C., Tan, G.W. &Welge, M.E. (2001). Knowledge management and data mining for marketing.Elsevier.*Decision Support Systems.* 31, 127-137.

Tse, G., Jason, T.& Patricia, A. (2010). Evolution Strategies for Evolving Artificial Neural Networks in an Arcade Game. Proceeding of Knowledge Management 5th International Conference, May, 2010. UUMCAS, Malaysia.

Ulrike, S. & Dorothy, L. (2002).Studying Knowledge Management in Information Systems Research: Discourses and Theoretical Assumptions.*MIS Quarterly.* 26(3), 213-242. Retrieved from: http://mmlab.ceid.upatras.gr/courses/AIS_SITE/files/projects2004/paper711/12_KM_in_IS_MISQ.pdf

Ward, J &Peppard, J. (2002). Strategic Planning for Information Sytems, 4th Edition. UK. John Wiley & Sons, Ltd. ISBN: 978-0470-84147-1

Yi, P., Gang, K., Young, Shi. &Zhengxin, C.(2006). A descriptive framework for the field of Data Mining and Knowledge Discovery.*International Journal of Information Technology & Decision Making.* 7(4), 639-682.

Young, R. (2010). Knowledge Management Tools and Techniques Manuals. United Kingdom. Asian Productivity Organisation. ISBN: 92-883-7093-7. Retrieved from: http://www.apo-tokyo.org/publications/files/ind-43-km_tt-2010.pdf

Zhan, J. (2008). Privacy-preserving collaborative data mining.*IEEE Magazine of Computational Intelligence.* 3(2), 31-41.

Zhao, F. (2001). Post Research Supervision: A Process of Knowledge Management. Centre for Management Quality Research, Bundoora West, RMIT Univeristy. Retrieved from: http://ultibase.rmit.edu.au/Articles/may01/zhao1.pdf