

Semantic Knowledge Retrieval KMS Components- Validating The Questionnaire Items

Abdulmajid B. Umar^{1,2*}, Azmi Jaafar¹, Marzanah A. Jabar¹, Masrah Azrifa Azmi Murad¹

¹Department of Software Engineering and Information Systems, Faculty of Computer Science and Information Technology, University Putra Malaysia UPM, Malaysia

²Department of Computer Science, Faculty of Science, Northwest University Kano, Nigeria

*Corresponding author email: bumar05@yahoo.com

Abstract: Knowledge management system (KMS) provides access to stored knowledge through the utilization of information and/or knowledge retrieval technologies. This enables knowledge workers to have necessary information for performing their knowledge intensive tasks effectively. To model the requirements for a semantic-based knowledge retrieval (KR) of such systems, a literature review of KMS (including the ontology-based) and semantic knowledge retrieval frameworks is conducted. By so doing, tools, technologies, and knowledge components necessary for a semantic knowledge retrieval KMS are identified and modeled as proposed. Questionnaire survey is used to verify the model components. Pilot survey is initially administered to validate the questionnaire items and to ensure that the item constructs are acceptable for determining the important components of the model. Rasch measurement is used for analysis due to the ordinal nature and collected small sample of the Likert-scale data. Questions perceived to be misleading are revised and some confusing or rather controversial questions are excluded from the questionnaire.

Keywords: knowledge management systems, semantic technologies, knowledge retrieval components, Rasch analysis, semantic search.

1. Introduction

Today, businesses are faced with a stark reality: “anticipate, respond, and react to the growing demands of the marketplace, or perish”[1]. Intense competitive environment demands innovative ways for the survival and actualization of business success. This results in the realization of the importance of knowledge to organizational sustenance. As it is now, only a few would debate validity of the statement that the world is “knowledge-based”. Researchers, academicians and industries therefore consider knowledge as a valuable resource that contributes to sustainable competitive advantage, thereby stressing the need to understand its multifaceted concept and multilayered meanings. In line with this understanding, Sarker [2] argued that the development of best practices for managing this complex concept of knowledge is the key to riding on today’s “competitive wave”.

Knowledge management has emerged as a major research topic within the past decade or two, mainly to study the effective management of knowledge. Notable definitions of this initiative appeared early [1-4]. Accordingly, knowledge management systems (KMS) surfaced as IT tools support the KM strategies implemented by organizations [5]. While early research for KMS models and frameworks emphasized on aligning the whole system with culture and people,

emergent views started to disintegrate the processes involved to provide closer insight into individual aspects. This growing trend of models for KMS led to knowledge creation, knowledge storage and retrieval, knowledge transfer, and knowledge share as very active sub-disciplines of knowledge management systems.

Knowledge retrieval in this context deals with identifying and leveraging technologies that facilitate access to information from KMS. If not properly studied, critical knowledge remains trapped in organizational knowledge assets that are not utilized by knowledge workers. Several different models of KM and KMS have been developed, with some focusing on particular sub-discipline while others provide a more generic view. Though knowledge retrieval has been a topic of recognition in this direction due to its importance, reviewed literatures lack generic formal model or framework for this sub-discipline within the KMS context. As such, we initiated a research on building a formal model. To identify important components for the model, a questionnaire survey is designed for onward distribution to researchers in Malaysian higher institutions. However, construct validity of the survey item needs to be checked. Thus, this paper aims to address this issue. We administered a pilot survey, which will be used to analyze the reliability of the questionnaire items. Rasch measurement model will be used for this purpose.

The paper is structured as follows: Section 2 discusses related research on formulating a semantic KMS model and validating framework construct. Methodology for this research is presented in Section 3, while results and discussion are in Section 4. Finally, Section 5 discusses the conclusion and future research work.

2. Related Works

2.1 Semantic and Ontology based KMS Frameworks

A number of studies were previously performed on formulating semantic or rather ontology-based KMS frameworks. In one study [6], an ontology-based KMS for flow and water quality modeling was presented, mainly to facilitate knowledge acquisition. The framework consists of application, description and object levels for interfaces, Ontologies, and knowledge objects respectively. The authors claim to provide intelligent knowledge access, although a critical examination of the framework does not indicate any component to support such claim. A more comprehensive framework of ontology-based knowledge management

systems which includes user, knowledge retrieval, knowledge sharing and reuse, ontology editing, and knowledge acquisition modules is described in another study [7]. Detailed aspects of the knowledge retrieval module were shown, which may lead to assumption that a certain level of semantic and intelligent knowledge access can be achieved from the framework. However, implementation and evaluation of the framework was not described, hindering a conclusion of the effectiveness of the framework. There are other studies of this nature [8-11]. More recently, a study [12] presents a model for semantic KMS.

On a more general note, KMS has been a topic of discussion in several research studies. Nevertheless, some limitations still exist from these studies. First, most of the frameworks are developed to handle specific organizational setting. Second, the semantic or ontology based KMS do not fully take advantage of semantic technologies as highlighted in one study [13]. They only tend to use Ontologies for knowledge representation. Additionally, most of the studies do not empirically prove the important components that make up the frameworks, except for a few (e.g. [12]) which suffer from the second limitation highlighted, in addition to the preliminary nature of the study.

2.2 Validating Framework Construct

Question on the validity of models and frameworks has been a lingering topic. While some researchers choose methods such as case study, others prefer surveys to identify relevant components of their models. Several validity checking methods have been used by different studies such as content validity, which is the degree to which the measure satisfactorily samples the content. Criterion validity is another method that checks the extent a measure empirically is associated with relevant criterion variables [14].

Construct validity is considered as a more general form of validity that entails the two described validity measures, and is the chosen method for this study. It refers to the extent to which a test measures what is supposed to measure [14]. Recently, there has been an increase in the number of researchers using the Rasch measurement model to check items' construct validity [12,15]. Though this model traditionally deals with competency evaluation of persons and items, its usage has been practically extended [16]. Hence, we chose to use similar approach to validate the construct of our survey item.

3. Methodology

This study was conducted using a literature review and questionnaire survey. To identify important components and factors for our study, a review of existing literatures on semantic knowledge retrieval, knowledge management systems, and related technologies was conducted. A questionnaire item was constructed to validate the components. 3 researchers, 1 from semantic information retrieval and 2 from the knowledge management specialization were given the draft questionnaire for possible additions or removal of questions. Before performing the actual data collection, a pilot study was conducted to test the questionnaire items.

Google forms were used to design and draft the questionnaire item. The survey was administered mostly through emailing of the link to researchers within the Malaysian public and private higher institutions. Excel was used to transform the collected data for onward loading to statistical analysis tool. Analysis was mainly conducted using Rasch measurement model. This model was selected due to its ability to handle ordinal Likert scale data, small sample size, and missing data. In addition, the model enables constructive analyses for persons and item statistics. It also allows for checking internal consistencies, reliability of response, and identification of outliers. Figure 1 depicts the research Methodology used.

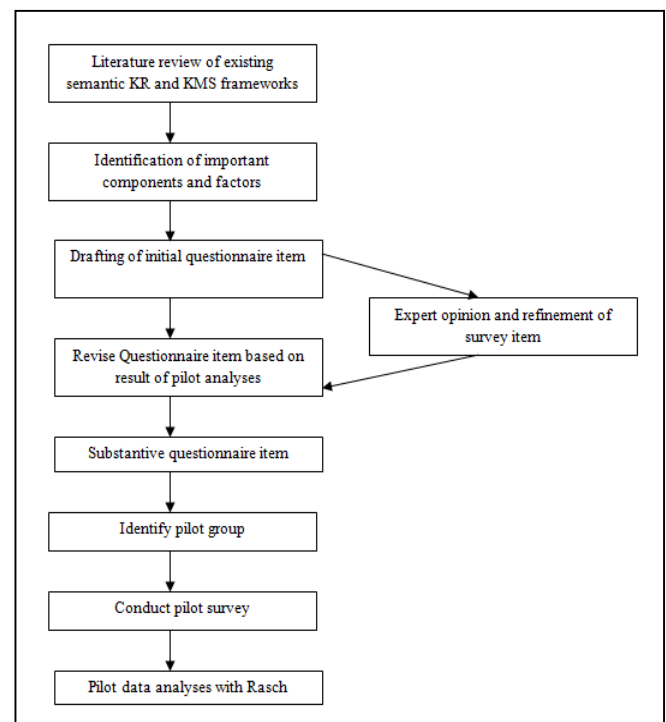


Figure 1. Methodology

4. Results and Discussion

The result of our review for existing semantic-based KMS and semantic knowledge retrieval frameworks is a synthesis of important components to be included in the proposed model. Table 1 is a summary of the findings.

30 respondents randomly drawn from public and private higher institutions in Malaysia participated in the pilot survey. This sample includes 7 academic staff, 17 postgraduate students (i.e. Master or PhD), and 6 categorized as 'others' entailing non-academic staff researchers and other category of postgraduate researchers. Collected data was tabulated and the response numbered from 1 representing 'strongly disagree' or 'not important at all' to 5 representing 'strongly agree' or 'very important'. To control participants' response pattern, some questions were negatively phrased. However, the response was transformed accordingly before the analysis was conducted. Winsteps is the Rasch tool used for the data analysis. The rating scale instrument quality criteria used is shown in Figure 2.

Table 1. Summary of Semantic Knowledge retrieval KMS components

Component	Factors	Sources
Semantic Interface	NL querying, Keyword querying, Intelligent result presentation	[8,13,17-22]
Semantic Processing	Similarity, Expansion	[7,13,17], [22-24]
Semantic Databases	Domain ontology, WordNet	[7-10,13,17,19,21, 22,24]
KM Activities	Knowledge generation, Knowledge search and retrieval, Knowledge share	[8,10-12]
Required Knowledge	Know-how, Know-what, Know-why	[12,25]
Functionality	Browsing, Expert finding, Messaging, News, Forum	[25,26]
Technologies	Computer-mediated collaboration, search engine, email, administrative tools	[9,25]
Knowledge Discovery	Knowledge push, Knowledge pull, Personalization	[9,12]
Other enabling features	Interoperability, Knowledge representation, Ease of use, Integration, Scalability	[13]

Based on the summary statistics of 30 measured persons as shown in Figure 3, none of the respondents was found to be extreme. Though respondents are not our target in this study, it is important to note that person reliability score of 0.83 obtained is quite good. This indicates that the responses are reliable for further analysis. Also, internal consistency of the item is reasonably good having obtained a Cronbach alpha value of 0.87.

Meanwhile, Figures 4 and 5 show the summary and measure order statistics of the item respectively. Infit (0.95 & -0.2) and outfit (0.98 & -0.2) values obtained for both mean square and z-standardized fall within the acceptable range, indicating fitness of data to the Rasch model. Accordingly, item reliability of 0.91 indicates that the data is excellent. Spread of data is $1.23 - (-1.67) = 2.90$ logit which is rather fair. This suggests that some items maybe poorly constructed.

Criterion	Poor	Fair	Good	Very Good	Excellent
Targeting *	> 2 errors	1-2 errors	< 1 error	< .5 error	< .25 error
Item Model Fit Mean-Square Range Extremes	< .33 - >3.0	.34 - 2.9	.5 - 2.0	.71 - 1.4	.77 - 1.3
Person and Item Measurement Reliability	<.67	.67-.80	.81-.90	.91-.94	>.94
Person and Item Strata Separated	2 or less	2-3	3-4	4-5	>5
Variance in data explained by measures	<50%	50-60%	60-70%	70-80%	>80%
Unexplained variance in contrasts 1-5 of PCA of residuals	>15%	10-15%	5-10%	3-5%	<3%
For Person/Item	Acceptable Range				
Acceptable Point Measure Correlation	0.4 < x < 0.8				
Infit/Outfit Mean Square	0.5 < x < 1.5				
Infit/Outfit z-standardized value	-2.0 < x < 2.0				

Figure 2. Rating scale instrument criteria [27]

SUMMARY OF 30 MEASURED Persons									
	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT MNSQ	ZSTD	OUTFIT MNSQ	ZSTD	
MEAN	110.5	27.8	1.13	.24	.93	-.4	.97	-.2	
S.D.	12.9	.5	.63	.05	.45	1.6	.49	1.5	
MAX.	132.0	28.0	2.69	.40	2.16	2.9	2.41	3.3	
MIN.	83.0	26.0	.08	.19	.28	-4.1	.35	-3.0	
REAL RMSE	.26	ADJ.SD	.58	SEPARATION	2.24	Person RELIABILITY		.83	
MODEL RMSE	.24	ADJ.SD	.58	SEPARATION	2.41	Person RELIABILITY		.85	
S.E. OF Person MEAN	= .12								
VALID RESPONSES: 99.4%									
Person RAW SCORE-TO-MEASURE CORRELATION = .97 (approximate due to missing data)									
CRONBACH ALPHA (KR-20) Person RAW SCORE RELIABILITY = .87 (approximate due to missing data)									

Figure 3. Summary of 30 meas

SUMMARY OF 28 MEASURED Items									
	RAW SCORE	COUNT	MEASURE	MODEL ERROR	INFIT		OUTFIT		
					MNSQ	ZSTD	MNSQ	ZSTD	
MEAN	118.4	29.8	.00	.25	.95	-.2	.98	-.2	
S.D.	17.7	.4	.92	.08	.36	1.1	.39	1.1	
MAX.	143.0	30.0	1.23	.42	2.19	3.4	2.23	3.2	
MIN.	90.0	29.0	-1.67	.17	.49	-2.0	.57	-1.9	
REAL RMSE	.27	ADJ.SD	.88	SEPARATION	3.21	Item	RELIABILITY	.91	
MODEL RMSE	.26	ADJ.SD	.88	SEPARATION	3.40	Item	RELIABILITY	.92	
S.E. OF Item MEAN = .18									
UMEAN=.000 USCALE=1.000									
Item RAW SCORE-TO-MEASURE CORRELATION = -.98 (approximate due to missing data)									
835 DATA POINTS. APPROXIMATE LOG-LIKELIHOOD CHI-SQUARE: 1750.45									

Figure 4. Summary of 28 measured items

For individual items however, point measure correlation is used to understand the response pattern. Almost all obtained results for this measure are of positive values, except for items I16 and I18. Nonetheless, those values in red box (i.e. I13, I22, I11, I12, I27, I27, I25, I04, I08, I01, I28, I03, I20, I26, 02) and the two negative values (I16 & I18) needed closer observation since they fall out of the quality criteria we are using in this study. Further analysis shows only items I13 and I12 of those items under observation failed to satisfy all Rasch critical fitness measures. Item I12 is a negatively phrased question of "having a traditional result list rather than intelligent result presentation". Item I13 is also a negatively phrased question that checks the agreeableness of "the system to understand user search intention and expand the query with synonyms for a more complete result". Since the importance of these items could not be deduced, the

questions are deleted from the questionnaire but replaced with better worded questions that use simpler terms to check the same factors.

ENTRY NUMBER	RAW SCORE	COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	OUTFIT MNSQ	PTMEA ZSTD	CORR.	ITr.	
24	90	30	1.23	.17	.99	.0	.96	-.1	.57	IO
23	93	30	1.14	.17	.80	-.9	.76	-1.0	.70	IO
15	94	30	1.11	.17	.59	-2.0	.59	-1.9	.75	IO
21	94	30	1.11	.17	.75	-1.1	.74	-1.1	.54	IO
19	98	30	.99	.17	.91	-.3	.86	-.5	.72	IO
14	100	30	.93	.18	.87	-.5	.83	-.6	.67	IO
6	102	30	.86	.18	.97	-.1	1.08	.4	.60	IO
17	101	29	.82	.18	.72	-1.2	.64	-1.4	.59	IO
9	105	30	.77	.18	.68	-1.4	.66	-1.3	.74	IO
10	106	30	.73	.18	.75	-1.0	.72	-1.0	.70	IO
5	107	30	.70	.18	.79	-.8	.89	-.3	.68	IO
7	107	30	.70	.18	.82	-.7	.77	-.8	.76	IO
13	115	30	.41	.20	2.19	3.4	2.23	3.21	.02	IO
22	119	29	.08	.23	.80	-.6	.72	-.9	.25	IO
16	120	29	.05	.23	1.34	1.1	1.34	1.1	-.04	IO
18	128	30	-.20	.24	1.10	.4	1.19	.7	-.01	IO
11	131	30	-.40	.26	.49	-1.7	.57	-1.4	.33	IO
12	128	29	-.52	.28	1.65	1.6	1.89	2.2	.18	IO
27	129	29	-.54	.29	.78	-.5	.84	-.4	.22	IO
25	133	30	-.54	.28	.81	-.4	.99	.1	.14	IO
4	135	30	-.71	.30	1.24	.7	1.24	.8	.19	IO
8	137	30	-.89	.32	1.59	1.5	1.68	1.7	.12	IO
1	139	30	-1.11	.34	1.03	.2	1.05	.3	.23	IO
28	139	30	-1.11	.34	.81	-.4	.90	-.2	.19	IO
3	140	30	-1.23	.36	.92	-1.1	1.08	.3	.04	IO
20	141	30	-1.36	.37	.83	-.4	.79	-.4	.29	IO
26	141	30	-1.36	.37	.67	-.9	.67	-.8	.48	IO
2	143	30	-1.67	.42	.74	-.6	.61	-.9	.49	IO
MEAN	118.4	29.8	.00	.25	.95	-.2	.98	-.2		

Figure 5. Item measure

Person P28 located at the top of the person distribution in Figure 4 is not considered extreme in response but significantly different in terms of the answer pattern. This person is an academic staff with knowledge of KMS's. It is possible that the person finds this study of much interest to them, hence the pattern of answers.

Also from the person-item distribution MAP in Figure 4, all the questions are not so difficult to understand since many items fall below the minimum person in the distribution. Item I02 is at the far end of the item distribution, indicating an overwhelming agreeableness to the question. The item is about the importance of "Knowledge storage and retrieval" among knowledge management activities. The response may suggest an empirical evidence of knowledge storage and retrieval to be the most important activity or process of knowledge management. However, this is not within the scope of this work and therefore left for future study. The item distribution seems quite closely bunched together except for items I13, I16 and I22. I13 is already deleted and replaced with a different question haven failed all critical measures. Participants may not have been clear on the term 'messaging' used in question I16 and 'administrative tools' used in I22. Thus the questions are revised and rephrased using clear layman terms for easier understanding. In all, only questions I12 and I13 are removed from the questionnaire while I16 and I22 are revised.

5. Conclusion

In knowledge management systems, knowledge retrieval is critical to ensure that organizational knowledge assets are not left un-used by knowledge workers. Furthermore, having adequate and precise knowledge when needed will also

ensure that valuable time is not spent searching for relevant information. Semantic knowledge retrieval makes this possible. To formulate a semantic knowledge retrieval KMS framework, the important components and factors of semantic knowledge retrieval, KMS and related technologies were synthesized from various sources. This was used to construct a questionnaire, which was tested through a pilot survey. The collected data was analyzed using Rasch to determine the items' construct validity as a measurement tool. Person and item reliability is found to be good and Cronbach's alpha was also within the acceptable range. The survey item is seen to fit the Rasch measurement model, though few items were found to be misfits. Questions perceived to be misleading are revised and some confusing or rather controversial questions are excluded from the questionnaire. It is our hope that the revision on questionnaire items will improve the reliability and construct validity of the survey items. The revised questionnaire will be sent out to several Malaysian higher institutions for further verification of the components for our proposed semantic KR KMS framework. With a substantive framework, a tool will be developed to help knowledge workers get adequate access to organizational knowledge assets.

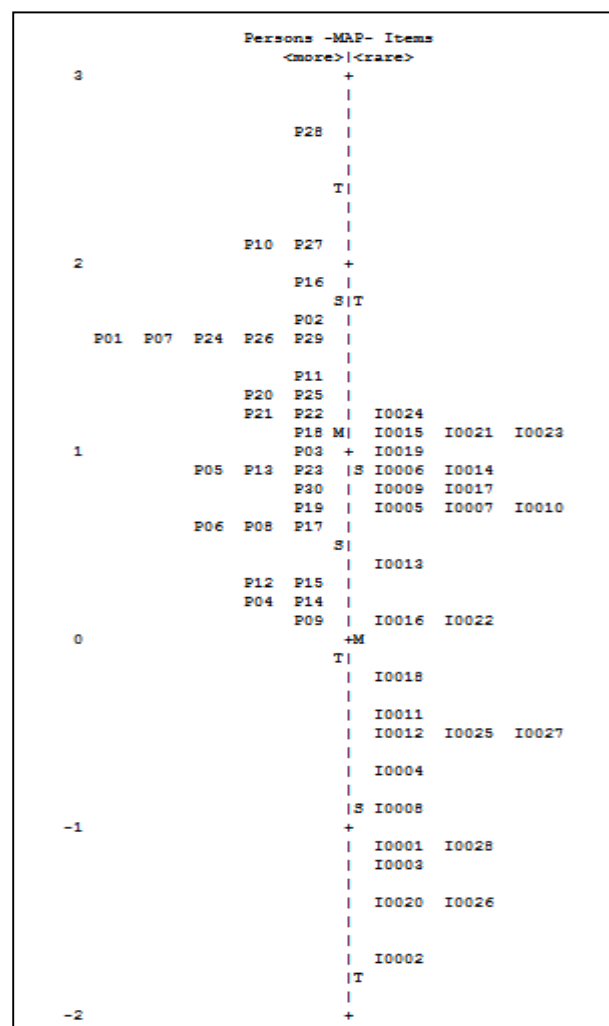


Figure 6. Person-Item Differential MAP

References

- [1] F. F.-H. Nah, J. L.-S. Lau, J. Kuang, "Critical factors for successful implementation of enterprise systems", *Business Process Management Journal*, vol. 7, no. 3, pp. 285–296, 2001.
- [2] S. Sarker, S. Sarker, D. B. Nicholson, and K. D. Joshi, "Knowledge transfer in virtual systems development teams: An exploratory study of four key enablers", *IEEE Transactions on Professional Communication*, vol. 48, no. 2, pp. 201–218, 2005.
- [3] T. H. Davenport, "Saving IT's Soul: Human-Centered Information Management", in *Harvard Business Review*, vol. 72, no. 2, pp. 119–131, 1994.
- [4] L. Prusak, "Where did knowledge management come", *IBM Systems Journal*, vol. 40, no. 4, pp. 1002–1007, 2001.
- [5] M. Alavi, D. E. Leidner, "Knowledge management and knowledge management systems: Conceptual foundations and research issues", *MIS Quarterly*, vol. 25, no. 1, pp. 107–136, 2001.
- [6] K. W. Chau, "An ontology-based knowledge management system for flow and water quality modeling", *Advances in Engineering Software*, vol. 38, no. 3, pp. 172–181, 2007.
- [7] H. Li, W. Li, Q. Cai, H. Liu, "A framework of ontology-based knowledge management system", *2nd International Conference on Computer Science and Information Technology 2009 (ICCSIT 2009)*, pp. 374–377, 2009.
- [8] E. Tello-Leal, A. B. Rios-Alvarado, A. Diaz-Manriquez, "A Semantic Knowledge Management System for Government Repositories", *2015 26th Int. Work. Database Expert Syst. Appl.*, pp. 168–172, 2015.
- [9] J. Zhang, W. Zhao, G. Xie, H. Chen, "Ontology- Based Knowledge Management System and Application", *Procedia Engineering*, vol. 15, pp. 1021–1029, 2011.
- [10] H. Faris, S. Totaro, A. Corallo, "Framework and implementation of a knowledge management system for aerospace collaborative working environments", *Proc. - IEEE Int. Conf. Mob. Data Manag.*, vol. 2, pp. 92–97, 2011.
- [11] M. A. Ale, C. M. Toledo, O. Chiotti, M. R. Galli, "A conceptual model and technological support for organizational knowledge management", *Science of Computer Programming*, vol. 95, no. Part 1, pp. 73–92, 2014.
- [12] Z. Che Cob, R. Abdullah, H. Risidi, M. Z. Mohd Nor, "Preliminary study on semantic knowledge management model for collaborative learning", *ARNP: Journal of Engineering and Applied Sciences*, vol. 10, no. 2, pp. 442–450, 2015.
- [13] M. D. Lytras, R. Garcia, "Semantic Web applications: a framework for industry and business exploitation – What is needed for the adoption of the Semantic Web from the market and industry", *International Journal of Knowledge and Learning*, vol. 4, no. 1, p. 93, 2008.
- [14] D. Westen, R. Robert, "Quantifying Construct Validity: Two Simple Measures", *Journal of Personality and Social Psychology*, vol. 84, no. 3, pp. 608–618, 2003.
- [15] M. N. Zali, R. Abdullah, M. A. A. Murad, H. M. Selamat, "KMS COMPONENTS FOR COLLABORATIVE SOFTWARE MAINTENANCE - VALIDATING THE QUESTIONNAIRE ITEMS' CONSTRUCT", *Journal of Information Retrieval and Knowledge Management*, vol. 1, 2011.
- [16] A. Abd Aziz, A. Mohamed, N. Arshad, S. Zakaria, S. Masodi, "Evaluation of Information Professionals Competency Face Validity Test using Rasch", in *Proceedings of the 5th WSEAS/IASME international conference on Engineering education*, 2008.
- [17] L. Xinhua, Z. Xutang, L. Zhongkai, "A Domain Ontology-based Information Retrieval Approach for Technique Preparation", *Physics Procedia*, vol. 25, pp. 1582–1588, 2012.
- [18] L. Tang, X. Chen, "The study of semantic retrieval based on the ontology of Teaching Management", in *Procedia Engineering*, vol. 15, pp. 1555–1559, 2011.
- [19] M. Fernández, I. Cantador, V. López, D. Vallet, P. Castells, E. Motta, "Semantically enhanced Information Retrieval: An ontology-based approach", *Web Semant. Sci. Serv. Agents World Wide Web*, vol. 9, no. 4, pp. 434–452, 2011.
- [20] A. Martín, "Intelligent Search Engine to a Semantic Knowledge Retrieval in the Digital Repositories", *International Journal of Advanced Robotic Systems*, vol. 8, no. 1, pp. 67–76, 2015.
- [21] H. G. H. Gao, J. Z. J. Zhao, Q. Y. Q. Yin, J. W. J. Wang, "Ontology-based enterprise information retrieval model", *2013 IEEE International Conference on Grey Systems and Intelligent Services*, pp. 1326–1330, 2009.
- [22] J. X. Shen, "Ontology-Based Semantic Retrieval for Management Information System", *Applied Mechanics and Materials*, vol. 278–280, pp. 2069–2072, 2013.
- [23] Y. Yuehua, D. Junping, H. Bowei, "A Novel Ontology-Based Semantic Retrieval Model for Food Safety Domain", *Chinese Journal of Electronics*, vol. 22, no. 2, pp. 247–252, 2013.
- [24] C. Hung, C.-F. Tsai, S.-Y. Hung, C.-J. Ku, "OGIR: an ontology-based grid information retrieval framework", *Online Information Review*, vol. 36, no. 6, pp. 807–827, 2012.
- [25] P. M. Smith, P. Meso, "A resource-based view of organizational knowledge management systems", *Journal of Knowledge Management*, vol. 4, no. 3, pp. 224–234, 2000.
- [26] S. Li, H.-C. Hsieh, I.-W. Sun, "An Ontology-based Knowledge Management System for the Metal Industry", *Twelfth International World Wide Web Conference*, p. 24, 2003.
- [27] W. P. Fisher, "Rating Scale Quality Instrument Criteria", *Rasch Measurement Transactions*, vol. 21, no. 1, p. 1095, 2007.