

Validating the Knowledge Representation Models

Hepsiba Vivenkanandan¹, Justus Selwyn²

^{1,2}Department of Computer Science, John Brown University, Siloam Springs, AR, USA-72761

Abstract—Knowledge representation models that can be used in any intelligent system (IS) are being proposed and in usage over the past four decades, and still newer models are being designed for the evolving intelligent systems world. The knowledge representation models that were proposed in our earlier works are being applied empirically in semi-controlled intelligent systems to validate in four dimensions: representability, existential validity, performance and scalability. The results are so significant and are helpful in providing an insight into the build of the IS in terms of design, functional and storage. The future directions of this work would be apply there representation models in real-time intelligent systems and study their performance and scalability.

Keywords— Knowledge representation models, Intelligent Systems, Performance, Scalability.

I. INTRODUCTION

Having proposed the knowledge representation models that have covered knowledge acquisition, codification, design, storage and retrieval in our earlier works [3, 5] we are presenting how the proposed models are valid in the platforms of their functionality. We will present the research methodology we have carried out, the model validations, and measuring the knowledge transformations using the proposed models.

The formalization of these knowledge models for knowledgebase improvement can be done in many ways [8, 11], we have identified four criteria to validate the knowledge representation models that are presented in the earlier chapters. The four criteria of formalizations are:

- 1) Representability
- 2) Existential Validity
- 3) Performance
- 4) Scalability

These criteria are used as tools to verify and validate the knowledge models. These validations are carried out in one of the two separate phases: Preliminary and Main studies.

In the preliminary study, we conducted few experiments to prove that the validation criteria mentioned above are true empirically and continue to interpret the models are valid based on the identified criteria. Each knowledge representation models are valid as far as they are confined within a controlled empirical environment, in terms of their ability to implement, readability, and understandability. These are termed as external attributes of the models, and they hold good. However, there are also other attributes, termed as internal traits, which need to be validated for the proposed models [4, 10].

Some of these internal traits – representability, existential validity, performance and scalability – are addressed here. How the knowledge units (representation models) are reasoned out and are able to be representable, what is their scope in existence for a given scenario, how the Knowledgebase (KB) will exhibit performance when the entailment computations are done, and the scalability of the KB in terms of volume and computations. These validations are done with preliminary experiments and other mathematical validations, wherever necessary.

II. RESEARCH METHODOLOGY

We opted to follow a most basic approach for formalization for this research work. This work is more focused on proposing knowledge models for improving the knowledgebases and on convincing the knowledge workers and practitioners about the worthiness of each model. Persuading the knowledge users to agree and endorse these knowledge models in terms of its internal qualities is a hectic task, which needed much work on conducting small experiments on various types of KB environments. Moreover, the research community needs more works on knowledge model's formalization in order to accept them into the Knowledge Management (KM) community. Hence, we followed this research methodology where we discuss about the data collection and the environment setup for the experiments for each facet of validations in the preliminary study, and extended with this environment to our two case studies in main phase of this research work.

A. Experimental Setup

Relational Database schemas were collected from two Information Technology (IT) companies and two Engineering Educational Institutions. The structured forms of these data repositories were then converted to object-relational database (ORDB) models so as to favour manipulation of its data and convert them to KUs. Additionally, we created knowledgebase for MES which were used directly for this research purpose.

- For Acquisition: The employee databases from the two IT companies are taken, and a generalized structure of the database schema was arrived at. To make knowledge acquisition valid, the structured repositories were created, and using Content Parse Model (CPM) [5] some concepts were acquired. The knowledge repository consists of 47 class_types, 23 relational tables and procedures, functions and triggers associated with the relational schema. These concepts were then manipulated to formulate a knowledge unit. In addition to this, we also used a 12 MIDI files to acquire concepts from the music scores, extracting the necessary musical notations that are later represented as Knowledge-Balls in the knowledgebase for IS. These MIDI sources of structured repository served as inputs to acquire

concepts that are accessed and computed for providing relevant knowledge for given music scores.

- For representation: Three object relational database schemas and eight MIDI structured repositories are considered for this experimental work. The Object Relational (OR) schemas are taken from two different IT companies, and one educational institution. Their complexity varied as they deal with different variety of data. These three sample schemas house an average of 27 tables per schema, 12 to 17 class-types per tables, and 4 to 8 conceptual relation per class_type. The concepts acquired from the existing structured repositories are represented in OSM and CSM models and use these class_types and relational tables for storage. The relations among these class_types are also defined based on the identified actors and entities in this preliminary study
- For reasoning: After the representation of the acquired concepts, we used first-order logic (FOL) to represent the concepts by valid and unambiguous definition. Once these definitions are complete and consistent, we then attempted to define a class_type for the function and predicate symbols using Object Structured Model (OSM) and Concept Structured Model (CSM). A total of 42 concepts for reasoning were identified for first case study and 14 concepts for the second case study. This knowledge modelling using OSM and CSM in ORDB became easier because the reasoning part is well-defined. The FOL representations are stored as meta-KUs in knowledgebase for inferences and entailments.
- For KB Performance: With all these objects instantiated for the class_types, the knowledgebase is rich with concepts that are conceptually related with each other. We evaluated the performance of the KB while storage and retrieval of these concepts take place. The algorithms and the computations we did for storage and retrieval are tested for their validity empirically by measuring the transaction throughput, hits/sec, and other non-functional parameters.
- For KB Scalability: With the existing experimental environment, and controlled data/knowledge repositories, the KB performance and the results were convincing and significant. However, if the concepts, contents mapped to the concepts, conceptual relations, and object instantiations grow as the KB improves, then how the KB Systems will respond to the exploding growth? This is addressed by simulating the voluminous processing, and manipulation of the concepts in the KB, and assessing their performance in higher workload. Hence, we foresaw this scalability of the KB in the future, and hence addressed the storage and retrieval issues in the cloud platform too.

The translated, hybrid schemas in ORDB are very appropriate for this study that they have helped validation of our knowledge representation models.

B. Preliminary Work

The structured repositories we collected, some designed and some translated from relational database schema served one or the other purpose in this research. Though we started with doing a vertical study on knowledge models, knowledge

management processes and knowledge engineering activities that are in research and in practice, we got interested in bridging the two domains of knowledge world – KM and KE. While KM is all about processes and Knowledge Engineering (KE) about engineering computer based intelligent products that will enable KM, we have attempted to perceive how a tacit knowledge could be interpreted and represented as an explicit knowledge unit.

Thus, the work started with learning to interpret a tacit form of knowledge to an explicit representable knowledge form. We preferred FOL to interpret a concept and reason out its validity and completeness in existence. A concept that is expressed should be implementable in a knowledge-based system, and hence we attempted to design and define three knowledge representation models. Thus, our preliminary works with these knowledge models.

We came out with six knowledge representation models which were able to capture the concepts defined in FOL. However, after careful refinement, and simple experimental works we concluded to stick onto four models, one for knowledge acquisition, two for KU representation and one for K-Ball representation. These works included tacit-to-explicit conversion, model framework and definitions, experimental works and results.

C. Main Work

Having done with the framework, design and definition of knowledge representation models, we narrowed down to conduct specific experiments to validate the models and assess the performance of the knowledgebase in terms of its non-functional attributes. The experiments conducted in the preliminary work were scaled up to this phase, as we had sufficient volume of data repositories.

We opted to carry out this main work with repositories in ORDB from two IT companies and one Educational institution as case study 1, and we opted to choose the Music Expert System (MES), an intelligent music tool designed for research studies and experimental works as case study 2.

In Case Study 1, we designed a smart E-Learning system that would help the learners both from the institution and from learning organization to interact and get what they want to learn, all in a smarter way of searching, fetching, and delivering. The knowledgebase involved in this study was quite complex enough to handle OSM, CSM objects and in addition to this the contents of the learning materials are also part of this KB. In this way, the KB is divided into two: one having the rich set of learning materials in different formats – like doc, txt, ppt, pdf, tweets, posts, images, audio and video – and the other part handling the collection of instantiations of objects from the class_types. These object are conceptually related to other objects, and at the same time linked to the contents in the other part of the KB. The computation processes and the algorithms involved in retrieving them are developed to have a KB with better performance than their counterparts in relational information or database systems.

In Case Study 2, we had Music Expert System (MES), a software developed for composing musical notes and playing the music score, specifically for experimental research

purpose alone. The MES derives an intelligent formation of musical chords for the given progression of musical score. The musical notations in MES are represented as K-Balls and are represented as CFRM model. The music notations are parsed and stored as a concept, conceptually related to the other musical concepts and Keys. These concepts are stored in collection class_types within a representation framework [3, 5] We designed a different knowledgebase for this case study exclusively for MES and the performance of the KB while playback of the music score and formulation of chords for the given series of music notations are assessed.

These two case studies help us validate the Knowledge Models that we have proposed for improving the knowledgebase, empirically and courageously authenticate our research work on this area.

With this research methodology we sustained our works in each phase and did not lose the sting of the problem as the work took turns in different directions throughout the course of work. We were finally able to link all of the works and integrate them to present a framework of knowledge representation models which would be complete in the aspects of definition and implementation, and especially validating the internal qualities and the external qualities of the models.

III. VALIDATING THE MODELS

The model of a system has to consider the global set of data to test its consistency and reliability in its functionality and performance. The representation models are validated in four dimensions: Representability, Existential Validity and Performance, and Scalability. In this section we will deal with the four dimensions of validity, proving each of them experimentally.

A. Representability

The ability to represent an acquired concept, a knowledge unit or a knowledge-ball in an unambiguous logic and in a knowledgebase is termed as the ‘representability’ of the knowledge models. Both these representable presentations are unique and are required in order to formalize the knowledge models, in terms of its consistency in logical entailments from existing or learned concepts and its ability to implement using a data-modeling technique, here we have chosen object-relational data modeling. Hence the CPM, OSM, CSM and CFRM models are validated in two facets as follows:

a) Logical Representations:

Statement 1: For all structured and unstructured repositories, concept acquisition entails the knowledgebase (KB), and the logically entailed concepts are added to the KB.

Let us assume the knowledgebase (KB) is initiated with S, a set of clauses believed to be true for this instance of KB. This is written as:

$KB \leq S$, where $S = \{ C_1, C_2 \}$, a finite set of clauses.

Given $C_1 (\{ \rho_1 \})$ and $C_2 (\{ \rho_2 \})$, where ρ_1 and ρ_2 are literals from learning

If $C_i \leq \{ C_1 (C_2)$ where C_1 and C_2 are resolvents in the KB and C_i is of the form of S,

Then it concludes, C_i is added to S, implies $S' C_i$

Since $KB \leq S$, and $S' C_i$

Where results as $S = \{ c_1, c_2, \dots c_i, \dots c_n \}$

Thus $KB \leq C_i$

Statement 2: For all concepts in KB, there exists function and predicate symbols whose qualities and relations are contained in a KU or K-Ball.

It is a true fact that $KB \leq S$, and S is a set of concepts (or clauses). The statement 2 is presented as:

$$\forall y. \exists x. [(f(x) \wedge P(x)) \supset (Q(x) \wedge R(x))] \quad (1)$$

Where $f(x)$ is function symbols in x, $P(x)$ is predicate symbol of x, $Q(x)$ is qualities of x, and $R(x)$ is relations of x.

If function and predicate symbols are P, and qualities and relations are Q, then (1) can be written in FOL as:

$$\forall y \exists x (P \supset Q) \text{ whose equivalent CNF form is } [\neg P, Q].$$

Hence, $\neg [(f(x) \wedge P(x)), (Q(x) \wedge R(x))]$.

If $[\neg (f(x) \wedge P(x))]$ is C_1 and $[(Q(x) \wedge R(x))]$ is C_2

Then $[C_1 (C_2) \theta]$ implies $\{ \rho_1 \theta \wedge \rho_2 \theta \} ' S$, (2)

Where ρ_1 and ρ_2 are literals contained in C_1 and C_2 respectively and

θ is the variable assignment $\mu[x]$ of x in C_1 and C_2 .

Therefore, from (2), it is resolved that for all concepts, there exists f(x), P, Q, and R, contained in S, implying to $KB \leq S$.

Definitive Representations:

All the concepts that form part of a sentence are representable, both in logic language or in programming language. However, we would like to substantiate that the entailed concepts in KB also are instantiations of the defined class_types. Consider the statements:

John is brilliant

Inclefs hires brilliant and hardworking persons

From these statements, we'll be able to acquire the Man, Brilliant, Company, and Person as function symbols and HardWorking() as predicate symbols. If all these concepts are entailed, we'll have the resolvent clause as “John is brilliant and works with Inclefs”.

The existing concepts and its resolvents are ably represented using OSM and CSM as shown in fig. 1.

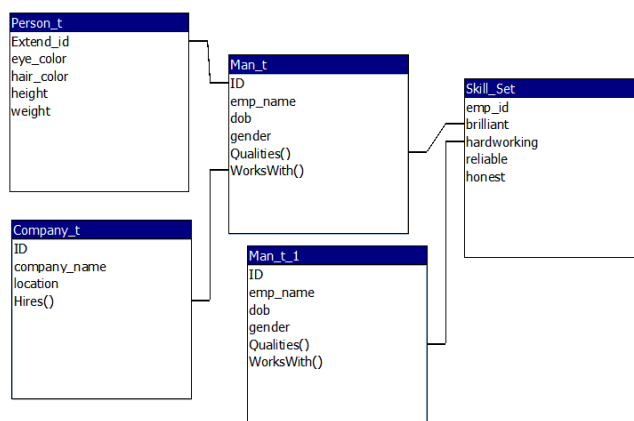


Fig. 1. Representations of Concepts in ORDB

The Person_t, Man_t, Company_t are defined as class_types using the OSM and CSM representation Models. The Person_t and Man_t class_types are modeled as OSM objects and they are existentially related, where Man_t inherits the Person_t. The qualities of Man_t Skill_Set:brilliant and Skill_Set:hardworking are in the concept (c) part of CSM bundling the Person_t, Man_t, and Skill_set and their relations with Company_t and its qualities are in the arc_set of the relation (r) part of the CSM.

B. Existential Validity

The four models for knowledge acquisition and representation are validated using Structured-Component-Connection (SCC) method [2, 6] approach. This approach is preferred for our research, because the models that we have proposed are subject to use by knowledge engineers who will create models or prototypes of the knowledge system, test and execute on a desktop, analyze for desired behavior and then scale the code to build a complete integrated knowledge system. Hence the prototypes of the models are validated and it can be scalable to real knowledge based system. As far as this research is concerned with building and improving knowledgebase, we have built a model of KB and tested it for its improved performance in our knowledge representation framework. Table I shows the details and specifications of KB built as a model.

TABLE I. Data Collection in SCC approach

Model	Structur es built	Compon ents integrate d	Connect ions	Front-end	Backen d
CPM	7	22	11	Java/C++	MySQL
OSM	15	46	24	Java/C++	MySQL
CSM	6	18	21	Java	MySQL
CFRM	12	29	17	Java/C++	MySQL
Knowledg ebase	2	8	12	Java	MySQL

To carry out this validation we have adopted the Structure-Component-Connection (SCC) approach [2]. This is a layered approach in which the entire model is built on its functional structure along with its behavioral components and in connection with other knowledge models. We choose “Educational System” as the basic ontology for this formalization approaches and improved it into knowledge based “Intelligent Educational System- KnowEdge”, which is the intended knowledgebase.

The proposed models are codified in Java and C++ language and the backend used to store the instances of these models is MySQL. The number of structures built in Java/C++ platforms, components integrated with the structure and the connections with the external structures are given in numbers. This is the environment set up where the proposed models are validated through experiments.

The models are tested for its reliability by generating structured data. These structured data are manually created, pulled in from legacy systems, and from simulated applications. The acquiring and representation of these data from the database are performed using our CPM, OSM and

CSM models. The processes of codification, representation and storing into the knowledgebase for the models are checked for its reliability. The knowledgebase we designed is reliable in its existence in a deployable platform. The knowledgebase schema we designed consists of class_types, tables, relations and objects instantiated corresponding to each model are given in Table II.

TABLE II. Experimental Setup for evaluating reliability

Model	Relational Tables	Class Types	Relations	Objects Instantiated (approx.)
CPM (acquisition)	4	28	3	300
OSM (Representation)	8	62	4	650
CSM (Representation)	2	11	16	130
CFRM (Storage)	5	20	4	200
Knowledgebase	12	174	22	1400

C. Performance

Performance evaluation of a system measures its behavioral complexity during runtime, under real world workload. Considering a knowledgebase as an individual system which stores, retrieves, manipulates and computes logical reasoning, we opted to evaluate its performance when the KB works on to receive persistent objects to be stored, accepts requests from KBS, perform computations, retrievals of the stored objects and service the requests from the KBS [1]. Hence assessing the KB’s performance and its improvement has become a mandate.

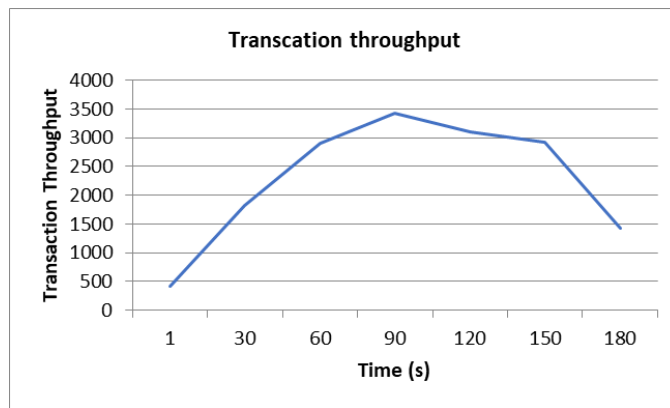


Fig. 2 (a). Transaction Throughput

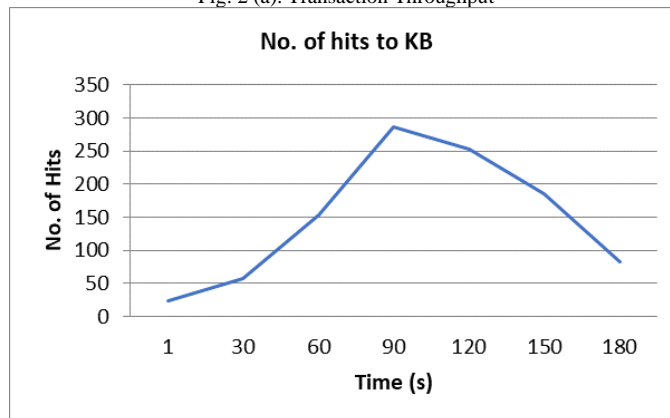


Fig. 2 (b). No. of Hits to the KB

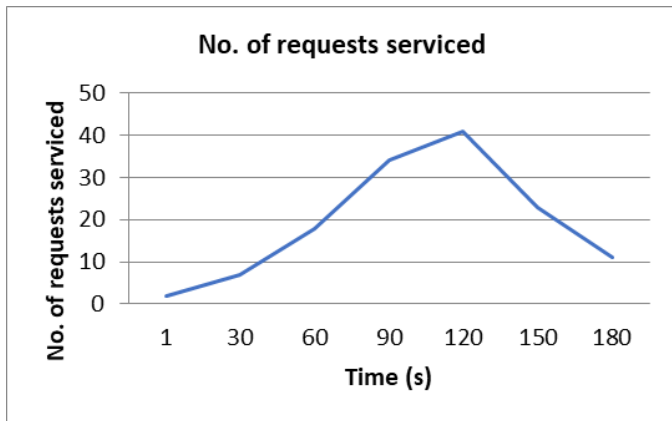


Fig. 2 (c). No. of Requests serviced by the KB

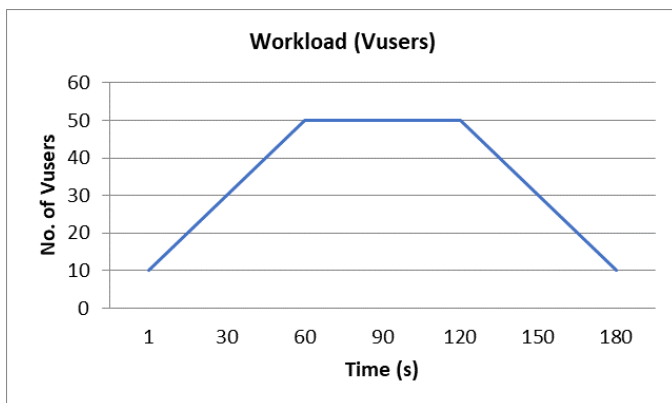


Fig. 2 (d). Workload given to the KB

While performance of a system covers various factors, we opted to choose performance transaction throughput, no. of hits to the KB, no. of requests serviced, all against a time frame of 1 – 180 secs in a given test scenario. Table III shows the collated performance test results.

These individual performance test evaluation reports show that the workload, shown in Fig. 2(d), in terms of the virtual users assignment to the KB is steadily ramped up for a period of 60secs and a steady workload is maintained for over another 60mins and then the ramped down over the next 60mins. The corresponding results for the transaction throughput are given in Fig. 2(a), which shows proportionality in the workload alone, and no bottlenecks are encountered in this controlled environment. The other two parameters of measures are shown in Fig. 2(b) and Fig. 2(c), all of them showing an equivalent proportionality to the number of Vusers in the Test scenario over a given period of time.

With this experimental setup given in Table II, the knowledge base is evaluated in terms of performance throughput of the knowledgebase. The relational tables, class_types, relations and objects instantiated are for an evaluation run for a duration of 180 secs (3 mins), for a parallel user load of 10, with an average of 3 transactions is carried out.

Fig. 3. shows the performance throughput results of the same test scenario of 50 Vusers executed over a period of 180 secs. The graph showed results as per the objects instantiated,

stored and fetched according to the ontology on which these models were designed.

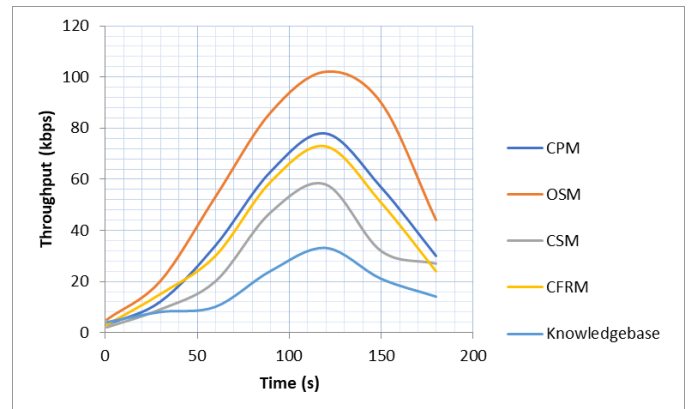


Fig. 3. Performance throughput of the knowledgebase

Table III shows the max, min and average of the performance throughput of the individual knowledge representation models and the knowledge base. This is the preliminary work we have carried out. More number of evaluations and validations are done in our main study.

TABLE III. Performance Throughput of the knowledgebase

Models	Performance Throughput in (kbps)		
	Max	Min	Average
CPM	78	3	39.57143
OSM	102	5	57.14286
CSM	58	2	27.85714
CFRM	73	3	36.42857
Knowledgebase	33	4	16.28571

D. Scalability

Scalability, in this context is defined as the extensibility of the design and definition of the proposed models and the knowledgebase itself. Hence, we evaluated scalability in three facets: design-level, functional-level and storage-level.

TABLE IV. Scalability at Design-Level

Models	Concepts		Relations		Objects	
	Prelim	Main	Prelim	Main	Prelim	Main
CPM	14	48	26	82	160	2000
OSM	22	86	47	180	3010	7000
CSM	20	82	70	320	2550	6300
CFRM	500	3200	1320	5400	4000	10200

a) Design-Level:

The CPM for knowledge acquisition involves the parsing of structured and unstructured forms of repositories. The parsing algorithms for structured repositories are applied for 20 contents in the beginning which we scaled up to 380 contents during our main study. The OSM and CSM models for knowledge representations are able to scale-up to accommodate any number of concepts and relations.

This is because we designed in such a way that the links are designed as pointers field, which links to other objects through address references. In CFRM model, obviously, the number of musical notations keeps varying based on the music score we take as input. The number of chords that can be entailed to the knowledgebase can be scaled up to take any

number of chords and notations. Table IV shows the scalability in our preliminary and main study, in the given controlled experimental setup.

Statically, the design of the models and knowledgebase can be scaled up in terms of concepts and relations, their acquisition and handling of instances of the class_types.

b) *Functional-Level:*

It is equally important to evaluate the functionality of the models in real-time empirical executions. Dynamic behavior of the objects as they scale up in terms of building up knowledgebase with contents, KUs and K-Balls, and the system adds-in knowledge through learning and logical entailments, is assessed. Behavior is assessed in terms of the transaction throughput of each of the contents of the KB. Fig. 4(a) shows the results of the throughput for the KUs and KBalls in the Part-A of the Knowledgebase, when these units are instantiated to populate from zero to 6000 units. Fig. 4(b) shows the results of the throughput of the contents (docs, audio, video, tweets and posts) in the Part-B of the Knowledgebase, when these contents are stored in the file storage system and accessed for manipulations as search results, reasoning and computations.

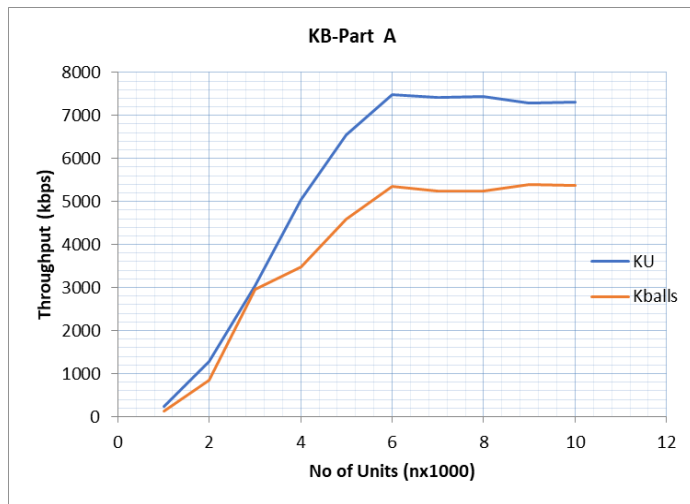


Fig. 4(a). Throughput for KU and KBalls in Part A

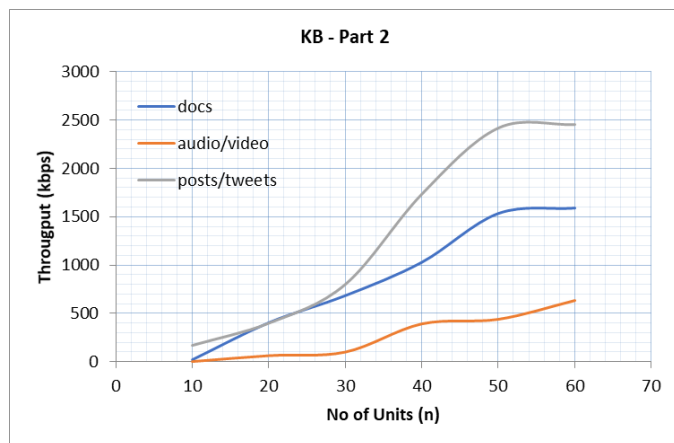


Fig. 4(b). Throughput for contents in Part B

From this assessment we were able to foresee that the KB when scaled up to an enormous volume, it will still be able to perform the required reasoning, computational searches and the required operations all in a better platform.

c) *Storage-Level:*

The validation of the KB in terms of scalability is substantiated in the two empirical studies done in our main studies. The project KnowEdge housed more than a thousand of KUs and K-Balls, and the objects of CFRM model has still more, because musical notes are instantiated into the collection object types rapidly, and their manipulation involved more combinations of musical notations.

We observed that the design level architecture holds good with our four knowledge models, and they are stable, adaptable and compatible with situations where KB shows linear scalability. In the sense, when users, contents and traffic increase proportionally, linear scalability is encountered. This is taken care by the underlying storage infrastructure of the RDBMS over which we are building our KB.

TABLE V. Scalability in Tera/PetaData Platforms

Model	EduNiversity	MES
Storage	144 1TB or 2TB SAS drives	8 300GB SSD per node
Total Capacity	Up to 186PB	Up to 18TB
Scalability	Up to 4,096 nodes	Up to 24 nodes
Memory	48GB per node	96GB per node
Workload	Analytical Archive	Very high-performance analytics

In this work, we built our KB over the RDBMS in MySQL for KnowEdge and Oracle for MES. The applications that are going to process the KB are of great concern that it has to handle the power of this voluminous KB. If the two case studies we are dealing with are scaled-up, then the specifications give in Table V need to be satisfied by the storage infrastructure.

In most situations, if the scalable (sustained) performance is measured against growing workloads in terms of volume and complexity, the Teradata platform architecture will definitely provide a more robust scale-up and scale-out model for the KB and its dependent applications.

IV. CONCLUSION

These validations of the knowledge representation models that may be used in any intelligent systems are carried out under semi-controlled environment. When run in real-time systems, the performances and the scalability of the models would face new challenges which are not addressed in this research work. The future progress of this work is directed towards conducting experiments with live robots or decision-making systems and study their performance with live data.

ACKNOWLEDGMENT

This work was partially done in the summer scholarship grant provided by John Brown University. We acknowledge the University's commitment in teaching and research, and providing infrastructure to carry out this research.

REFERENCES

- [1] 67. Dina M.A. and Christina C. H., "Building Decision Support Systems for Sustainable Transitions" in *Business Transitions: A Path to Sustainability*, Springer, pp. 239-249, ISBN 978-3-031-22244-3, 2023
- [2] Gartland-Jones. A., and P. Copley, "The Suitability of Genetic Algorithms for Musical Composition", *Contemporary Music Review*, Vol. 22, No. 3, pp. 43-55, 2003.
- [3] Hepsiba Mabel., Justus S., "Performance Assessment of 3-Layered Knowledgebase and Representation Models: An Empirical Study", *International Journal of Knowledge Engineering and Data Mining*, Vol. 3, No. 4, pp. 239-254, Dec 2015
- [4] Jiang, L.; Shi, J.; Wang, C. Multi-ontology fusion and rule development to facilitate automated code compliance checking using BIM and rule-based reasoning, *Advanced Engineering Informatics, International Journal*, Vol. 51, pp. 101-449, 2022
- [5] Justus Selwyn, "Design of Knowledgebase for EES using Object-Relational Data Modelling", *International Journal of Computer Applications*, Vol. 184, No. 21, 2022
- [6] Li, H., Yang, Z., Jin, C. and Wang, J. "How an industrial internet platform empowers the digital transformation of SMEs: theoretical mechanism and business model", *Journal of Knowledge Management*, Vol. 27 No. 1, pp. 105-120, 2023
- [7] Nanda J, Thevenot H, Simpson T, Stone R, Bohm M, Shooter S., "Product family design knowledge representation, aggregation, reuse, and analysis", *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, Vol. 21, No. 2, pp. 173-92, 2007
- [8] Pawel Garbacz, Piotr Kulicki & Robert Trypuz, A formal ontology of knowing and knowledge, *Knowledge Management Research & Practice*, Vol. 10, No. 3, pp. 206-226, 2012
- [9] Park, H.; Shin, S., A Proposal for Basic Formal Ontology for Knowledge Management in Building Information Modeling Domain, *Applied Sciences*, Vol. 13, pp. 4859, 2023.
- [10] Pierre B., Richard E.F., "Guide to the Software Engineering Body of Knowledge - SWEBOK", IEEE-Computer Society Press: www.swebok.org, Los Alamitos (USA), 2014.
- [11] Ronald J. Brachman and Hector J. Levesque, *Knowledge representation and reasoning*, 2nd edition, Elsevier publications, 2004.