https://doi.org/10.31436/ijpcc.v9i2.399

Semantic Similarity Exploration in Heterogeneous Sparse Multidimensional Numeric Spaces

A Case Study of the Quran Text using SemSim

Adel Sabour[1]*, Abdeltawab Hendawi[2] and Mohamed Ali[3]
[1,3]University of Washington, Tacoma, USA
[2]University of Rhode Island, USA

*Corresponding author: sabour@uw.edu
(Received: 3rd April 2023; Accepted: 18th May 2023; Published on-line: 28th July 2023)

Abstract— Comparing heterogeneous entities that seem to have no common denominator, and inferring similarities among their attributes is a complex process compared to homogeneous entity sets. Two types of similarities exist: semantic similarities and numeric similarities. This paper presents a system named SemSim that helps users explore similarities in heterogeneous environments of multidimensional data sets. The system helps the user (a) define entities, entity groups, and dimensions of entities, (b) detect numeric similarities among entities either across the same dimensions or across different dimensions, (c) correlate the semantic similarities given in a knowledge graph with the detected numeric similarities, and then (d) use the detected numeric similarities to enhance the knowledge graph by exploring and mining for other hidden semantic similarities. As a case study, we apply the proposed system to the text of the holy Quran to explore the correlation between the semantic similarities and the numeric similarities for chapters, verses, and words.

Keywords— Similarity detection, multidimensional data, heterogeneous entities, semantic similarity, numeric similarities, knowledge graphs, Quran, and text analysis.

I. Introduction

A car is an entity. The car, as an entity, has multiple attributes (or dimensions). Example dimensions include the engine size in cubic centimeters, the fuel consumption in miles per gallon, and the number of seats. All car entities (e.g. car makes and models) fall under an entity group that may be called the car entity group. Similarly, other entity groups can be defined to include other types of entities like airplanes, bikes, books, and movies, to name a few examples.

Numeric similarities can be defined and measured across the multidimensional space of each entity group. Cars can have similar engine sizes or fuel consumption rates in miles per gallon. There is a body of research that studies and introduces various algorithms to detect numeric similarities between points in a multidimensional space [1] using nearest neighbour queries [2], range queries [3] and clustering techniques [4].

Semantic similarities between entities (in contrast to numeric similarities) are human-judged or evaluated through various machine-learning techniques. Two cars may be identified as semantically similar in being popular car models. Two movies may be identified as semantically similar in being violent, comic, or drama. Two books can be identified as semantically similar because they both address the same issues.

So far, we have highlighted two types of similarities (numeric and semantic similarities) and we have introduced the concept of an entity group that represents a group of multi-dimensional entities of the same type (or the same dimensions). We have also introduced the possibility of comparing entities numerically or semantically as long as (a) we compare entities from the same entity group and (b) we compare entities along the same dimensions. For example, we compare two car models (as two entities in the same car entity group) along the mileage per gallon dimension (which is the same dimension for the two entities). In this research, we question the possibility of comparing two entities from different entity groups across the same/different dimensions, possibly after normalizing the numeric value [5]. For example, we may compare the mileage per gallon and the engine size of specific airplanes to the mileage per gallon and the engine size of cars, after normalizing the magnitude of the numeric values so that the comparison is feasible. Moreover, we question the value of comparing entities from different entity groups across different dimensions and correlating the numeric similarity of these entities (across different dimensions) to the semantic similarities between these entities. For example, two cars can be popular (which is an observable semantic similarity) although these cars have no similarities across the same numeric dimensions. These two cars are both popular because one of them is the

https://doi.org/10.31436/ijpcc.v9i2.399

most economic car (as measured by the mileage per gallon) and the other one is the most luxurious (as measured by the engine size). Therefore, the numeric similarity across different dimensions (i.e., having the highest values over the fuel consumption and engine size dimensions) made these two cars similar over the popularity semantic dimension.

Based on the example above, we believe that semantic similarities and numeric similarities are tied together. However, the numeric similarities that result in semantic similarities do not have to be along the same dimension. Numeric similarities across heterogeneous dimensions may be the cause of semantic similarity. A combination of numeric similarities across different dimensions may stand behind the human-observed semantic similarity. In this research, we bridge the gap between similarities in the numeric space and the semantic space. Guided by semantic similarities (as given by a knowledge graph), we build a system that helps the user explore semantic similarities between various combinations of numeric heterogeneous dimensions. We understand that there can be an enormous number of possible combinations of the numeric dimensions. We mine for the impactful combinations of numeric dimensions that stand behind semantic similarities. After we detect the impactful combinations of numeric dimensions, we use these dimension combinations to detect additional semantic similarities that may be hidden in the semantic space and, consequently, improve the knowledge graph.

As a case study, we use this system to detect the relationship between the semantic space and the numeric space using the holy Quran text. Chapters, verses, and words are entity groups. Each entity group has multiple numeric dimensions like the number of letters, number of words, number of verses in a chapter, the chapter serial number, the word serial number from the beginning of the Quran, the word number from the beginning of the verse or the beginning of the chapter. On the semantic level, chapters, verses, and words are associated with topics and, therefore, semantic similarities can be observed among these entities and entity groups. The proposed system helps researchers detect the correlation between the semantic similarities and the numeric similarities in the holy book of the Quran.

The contributions of this paper can be summarized as follows:

- We develop a model to define multidimensional entities and entity groups in an environment characterized by heterogeneity in dimensions.
- We build a system, called SemSim (for Semantic Similarity), that helps the user explore numeric similarities between entities that span different entity groups and, possibly, across different dimensions.
- Guided by a knowledge graph that represents the semantic similarities between entities, the system alters the user when a numeric similarity is found between entities that already show semantic similarities.
- The system detects the numeric dimensions that impact the semantic sim- ilarity and uses these impactful dimensions to further detect hidden se- mantic similarities
- We use the text of the holy Quran as a case study to explore the semantic and the numeric similarities over chapters, verses, and words.

The rest of this paper is organized as follows. Section 2 overviews related work. Section 3 presents our model for the multidimensional numeric space and the semantic space. Section 4 highlights the proposed system's architecture along with the system. The paper is concluded in Section 5.

II. RELATED WORK

In this section, we overview related work across three directions. First, we start with the research that tackles the semantic aspects of the Quran and, more specifically, the research approaches that build ontologies on the Quran text. Second, we overview some techniques that aim at discovering similarities in the Quran text using mathematical methods and clustering approaches. Finally, we highlight a few relevant papers that address heterogeneous multidimensional space.

Mohammad and Eric [6] surveyed thirteen Quran ontologies. This survey compared various ontologies across fourteen criteria. They presented the com- mon data sets that contain the Holy Quran. Based on the survey, the authors highlighted Arabic Quran Corpus (AQC), Quranic Topics (QT), and QuranA to be among the prominent ontologies of the Quran.

Ismail et al. [7] utilize NLP techniques to learn and build ontologies. In a sample of five chapters that contain 176 verses, the validity of the extracted synonyms and definition relations was 51.52%. In order to improve the result, the authors recommended that NLP patterns should be modified in a way that is compatible with the Quran. The work in [8] builds an ontology in collaboration with specialists in the Quran and its linguistic constructs in order to reach a high level of accuracy and reliability. Alshammeri et al. [9] presented sentence-level analysis through a deep learning approach using a paragraph vector model to help identify key concepts in the text and their associated topics. By representing the semantic structure as dimensions in space, these vectors are used as building blocks for topic analysis and clustering. It was used in a K-Means clustering algorithm to create groups of related verses. The authors compared the result with a tagged corpus to calculate the result's significance. Several research directions showed interest in detecting similarities between verses in the Quran.

https://doi.org/10.31436/ijpcc.v9i2.399

Huda et al. [10] measured similarities between verses of the Quran using cosine and jacquard similarity and correlation coefficients. The authors implemented the proposed model on Surah Al-Baqarah in English. The authors highlighted the difficulties in determining the theme of the verse based only on the words and discussed why the context of the verse must be taken into account to obtain better results.

Khadangi et al. [11] studied the similarity between concepts presented within the chapter and the title of the chapter through NLP techniques over the root words. Then, the authors evaluated and compared the result to a random arrangement of verses. The result showed that the similarity of the internal concepts of the chapters (as arranged in the Quran) is twelve times higher than a random arrangement of verses.

With a large group of entities and entity groups, we expect a large number of attributes or dimensions to be present in the system. Such a high number of dimensions leads to a 'curse of dimensionality [12]. There are several ways to solve such a high-dimensionality problem and to leverage the sys- tem to handle big data. Chee et al. [13] surveyed existing methods that address dimensionality reduction in heterogeneous environments. The survey compared nine methods of dimensionality reduction and presented their points of strength and weakness. The survey presented various methods of data reduction such as data decompression, de-duplication, dimensionality reduction, and network theory-based methods. The authors stated explicitly that this area needs more research attention to improve the results given that there is no single solution that can fully manage the problem.

III. The Proposed Model for the Numeric and Semantic Spaces in a Heterogeneous Environment

In this section, we define our notion of a heterogeneous numeric multidimensional space as well as the semantic space. We define entities, entity groups and dimensions of these entities. We provide a formalization for the similarity problem under investigation and its associated query types. This section provides a generic formalization of the problem while section 3.3 provides an explanation of the dataset used and its representation in numeric and semantic spaces.

A. The Heterogeneous Multidimensional Numeric Space

A heterogeneous domain is a domain that has multiple entities and entity groups [14]. For example, a heterogeneous domain includes entities from the *car* entity group as well as entities from the *aeroplane* entity group. All entities are multi-attribute or multidimensional. Each entity group is characterized by the same set of dimensions. All cars have the same dimensions, all aeroplanes have the same dimensions and, similarly, all entities coming from the same entity group have the same dimensions. However, dimensions are expected to be different across entity groups. Cars and aeroplanes may share several dimensions but not all dimensions. There are some attributes (or dimensions) that are unique and specific to cars or aeroplanes. Our proposed multidimensional numeric space, as given by Definition 1, is composed of the union of all dimensions in all entity groups. Therefore, entities are expected to have numeric values over dimensions that are defined within their entity group and have null values otherwise (or across dimensions that are not defined for the entity's entity group).

Definition 1 The numeric space *is composed of E, a set of K-dimensional entities that has n entities. There are also m overlapping entity groups (each group is a subset of E). Each entity e_i where $1 <= i <= n$ belongs to g entity groups where $0 <= g <= m$. Each entity e_i has k_i non-null dimensions and, consequently, $K - k_i$ null valued dimensions, where $k_i << K$. The number of non-null dimensions (k_i) is variable across entity groups. For each entity e_i, and for each non-null dimension k, the entity has a numeric value $v_{i,k}$.*

B. The Semantic Space

Consider a knowledge graph [15] $KG(E, W)$, where entities in E, the set of K-dimensional n entities, represent the vertices of the graph and $W = w_{i,j} | w_{i,j} = Sim(e_i, e_j)$ is the set of edge weights between entities.

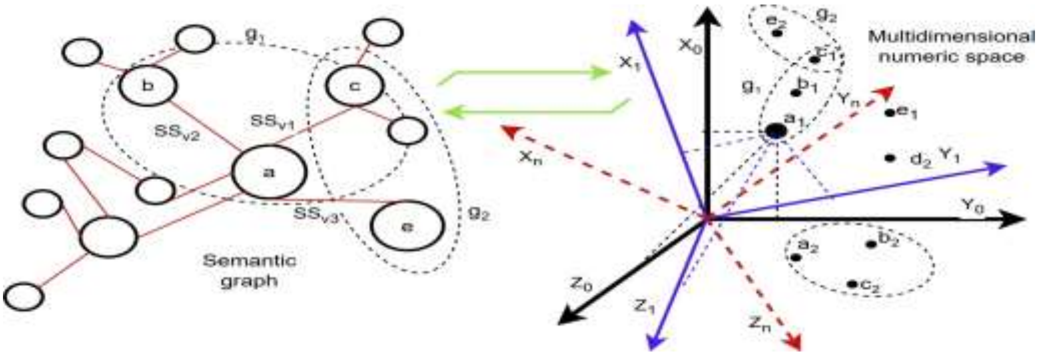

Fig. 1 Proposed model for the overlap between numeric and semantic similarities in a multidimensional space

https://doi.org/10.31436/ijpcc.v9i2.399

The weight $W_{i,j}$ represents a semantic similarity measure between entities e_i and e_j. Figure 1 presents the proposed integration between the numeric and semantic spaces. The left side of the figure presents a knowledge graph over entities, where a node (e_1, e_2, ...e_n) is an entity and an edge between two nodes is the semantic similarity weight between these two entities (SS_{v1}, SS_{v2}, SS_{vn}). These edges can be between entities (nodes) within the same entity group or they may extend across different groups (g_1, g_2, ...g_m). On the right side is the multidimensional space where each entity is represented as a point in a multidimensional space.

C. *Dataset Description and Characteristics*

According to Islamic belief, the Quran is the word of God revealed to Prophet Muhammad peace be upon him (PBUH). The purpose of our case study is to utilize the proposed system SemSim to discover and elicit knowledge from the Quranic text. Finding relationships and patterns between words, verses, and chapters. As shown in Table 1, the surah is an entity group that has verses as its entities. Example dimensions could be the number of words, the verse number, the surah number, and the number of characters per verse. Note that SemSim provides a user interface that gives the flexibility to the user to define entities, entity groups, dimensions, and dimension values dynamically and, then, to interactively explore the numeric and semantic similarities between entities, as described later in section 4. Using NLP techniques that analyse the context and the topics addressed in each surah or verse an ontology [16] can be constructed to capture the semantics of these entities. Moreover, a knowledge graph, with entities represented as vertices, can capture the semantic similarities among these entities. Such similarity scores become the edge weights of the graph to reflect the semantic relationships between the surahs, verses, or words. The dataset used in this study pertains to the Holy Quran. Specifically, the narration of Hafs. Hafs is the most widespread narration in the Islamic world. Hafs follows the Kufic verse counting scheme, which contains 6236 verses. Each verse is composed of multiple words, and each word is composed of Arabic characters. The verses are divided into 114 surahs.

TABLE I
AN EXAMPLE OF THE DISTRIBUTION OF THE PROBLEM ELEMENTS IN QURAN CASE STUDY

Numeric Space

Group (E)	Entity (e)	Dimension (k)	Value (v)
Surah	Verse	No. of words, verse no., No. of letters, etc	Value of k

Semantic Space

Vertices (e)	Edge (k)	Weight (w)
Concepts	Similarity	percentage of similarity

1) *Data Application: Exploring Semantic and Numeric Similarity*

For example, no one can know when he will die during his lifetime. We searched for the death and age of the Prophet Muhammad PBUH, in the Quranic text, and we know that he died at the age of 63. The word "Natawafynk" (meaning when you will die, Muhammad PBUH) was mentioned three times. For this case to be represented by a group for the word 'Natawafynk' and every entity is an instance of that word. With all the dimensions for each word. So we want to search for any dimension whose sum across all entities is equal to 63. We will notice that the sum of the dimension "surah number" across all entities is equal to 63. They are Surah No. 10, Surah No. 13, and Surah No. 40. It is from semantic compatibility. The Prophet Muhammad PBUH stayed in Medina for 10 years. And he stayed in Macca for 13 years. And he stayed 40 years before the message. This exactly matches the dimension of the surah number. We can get this result from the next equation.

$$(g_j, K, v)|K = k_1, k_2, ...k_m, f_x = 1^n v_k = v, k_d K \qquad (1).$$

For another example, the word" Raad" or "thunder" appears twice in the Qur'anic text, with the same wording and the same meaning. In Surah 2, verse 19, and Surah 13, verse 13. And when exploring the numerical similarity of the dimensions of the two words. It was found that the number of words used in the first and second verses is equal with a value of 19. It was found that the number of letters used in the first and second verses is equal to the value of 19. The number of repetitions of the letters of the word "Raad" in the first and second verses is equal to the value of 9. This example shows the numerical similarity associated with the existing semantic similarity. As a general template for all queries like this example. When there is a semantic similarity between e_i and e_j we get back all the equal values v across the same dimensions k. We may deal with the model in reverse. That is, when the value of the dimensions is similar across the same dimensions k, we display e_i and e_j, to discover a semantic similarity between them.

$$(e_i, e_j, k, v)|v = v_i, k = v_j, k \qquad (2)$$

oblem.

IV. THE PROPOSED SYSTEM

The SemSim system consists of three sub-systems designed to detect numeric and semantic similarities among entities in multidimensional datasets. These sub-systems are an ontology-building [17] sub-system (OBS), a multidimensional numeric representation [18] sub-system (MNRS), and a semantic-numeric similarity detection sub-system (SNSDS). In this section, we will present an overview of the proposed system's structure, followed by an explanation of each

https://doi.org/10.31436/ijpcc.v9i2.399

subsystem in the context of the entire system.

A. *The Proposed System Architecture*

Figure 2 depicts the system, its three subsystems, and the components of each subsystem.

The **MNRS** is a sub-system that processes numeric data in multidimensional datasets and has six components. Data Source Specification allows users to specify the data source. While Entity Processing creates and classifies entities. Multidimensional Processing constructs dimensions and processes entity values, and Dimension Construction constructs dimensions based on entity properties. Finally, Value Processing processes entity values to create a multidimensional representation.

The **OBS** builds and manages ontologies and shares two components with the MNRS (Data Source Specification and Entity Processing). It also has four components of its own. Ontology Editor allows users to create and modify ontologies, while Concept affinity ratio specification determines the strength of relationships between concepts. Knowledge Base stores and manages ontology- related knowledge.

The **SNSDS** detects similarities among entities in multidimensional datasets, attempts to find a relationship between the results of OBS and MNRS, and has five components. Entity Visualization provides a visual representation of entities and their relationships. Similarity Thresholding sets the similarity detection threshold, and Interactive Controls allow users to modify similarity settings. Finally, the Similarity Notification notifies users of similarities.

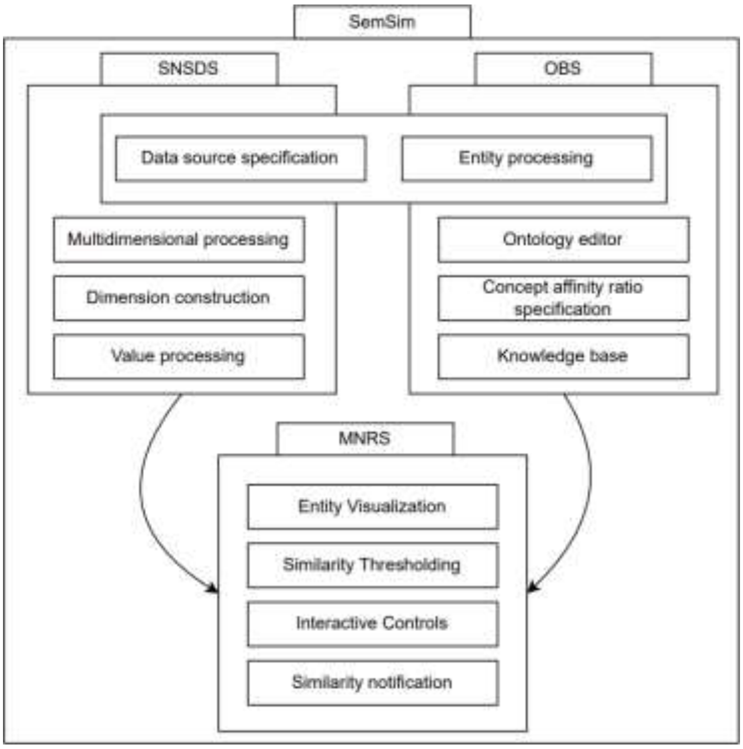

Fig 2. The Proposed System Architecture of SemSim: A System for Multi-dimensional Numeric and Semantic Similarity Detection

B. *The multidimensional numeric representation sub- system (MNRS)*

In this section, we discuss how the MNRS operates in two steps. The first step, which is entity processing, creates entities from a specified data source and classifies these entities into entity groups. The second step, which is the multidimensional processing, constructs the dimensions of the multidimensional space and processes the values of each entity across these dimensions. The following sections provide an overview of these two steps.

1) *Entity Processing*: Entity Processing has three sub-steps: a) identifying the data source, which is the underlying database(s) that the system will import the data from, as shown in Figure 3; b) creating entity groups that act as containers for the entities that are to be created in the following step; and c) creating the entities and assigning them to the entity groups. Entity groups and entities are created through SQL statements that query the underlying data sources, as shown in Figure 4. Figure 5 shows the created entities along with their entity groups and gives the user the ability to edit and delete entities and entity groups.

https://doi.org/10.31436/ijpcc.v9i2.399

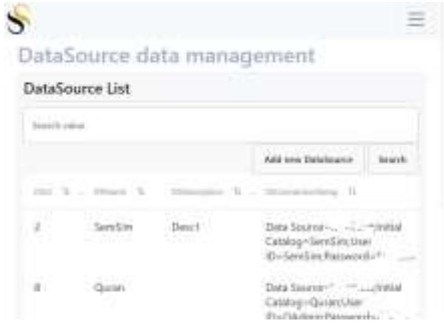

Fig. 3 Identifying the underlying data sources.

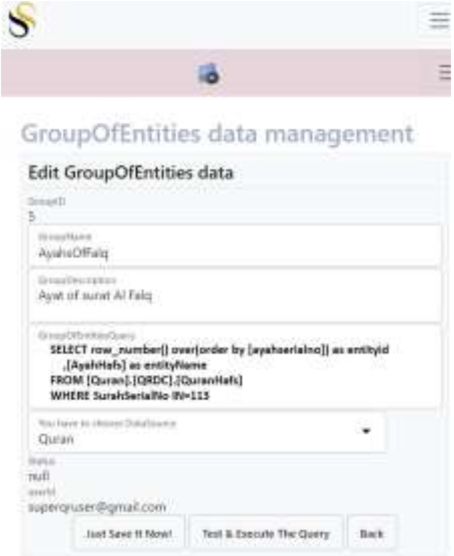

Fig. 4 Creating entity groups by querying the underlying data source

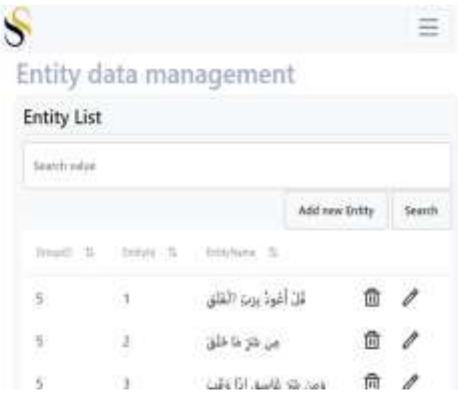

Fig. 5 Creating entities by querying the underlying data source.

2) *Multidimensional Processing:* As discussed earlier, the system accepts entity groups where each group has a set of dimensions. A dimension can be common to multiple entity groups or can be unique to one entity group. For example, the number of words dimension is common to the surah and the verse entity groups. However, the surah number is unique to the surah entity group. The multidimensional processing step includes three sub-steps: a) Defining the dimension list of all dimensions across all entity groups, as shown in Figure 6, b) assigning dimensions to entity groups, and c) querying the values of each. dimension in each entity from the underlying data source. Sub-steps (b) and (c) are done using SQL statements, as shown in Figure 7.

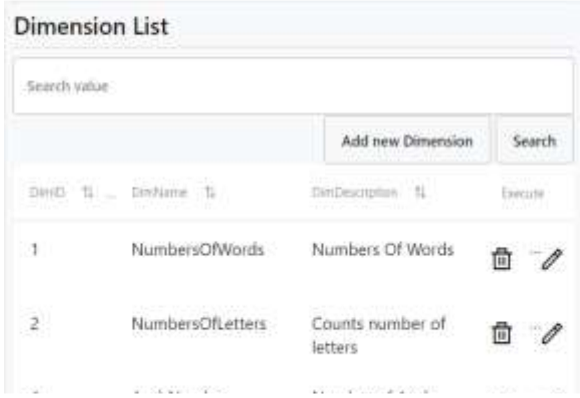

Fig. 6 Creating the list of expected dimensions

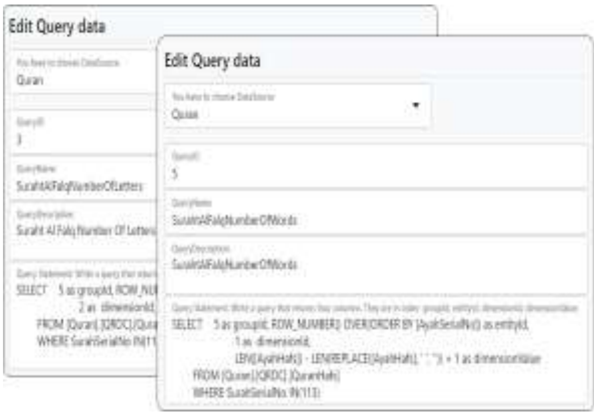

Fig. 7 Dimension Values Filling via SQL Statements

3) *Level-3 Heading:* A level-3 heading must be indented, in Italic and numbered with an Arabic numeral followed by a right parenthesis. The level-3 heading must end with a colon. The body of the level-3 section immediately follows the level-3 heading in the same paragraph. For example, this paragraph begins with a level-3 heading.

C. *The ontology-building sub-system (OBS)*

The Ontology Building System OBS module enables the user to specify the structure and relationships between different entities in the ontology. The system offers multiple ways for users to specify concept affinity ratios, including manual input, SQL statements, and the option to import existing ontologies. Furthermore, the OBS displays a list of relationships between entities along with their corresponding percentages of proximity, as shown in Figure 8.

https://doi.org/10.31436/ijpcc.v9i2.399

Fig. 8 List of Entity Relationships and Proportions in a Semantic Space

D. The Semantic-numeric similaritydetection sub-system (SNSDS)

SNSDS enables the user to explore relationships between entities across both the numeric and semantic spaces. Its visual representation of entities and their similarities elevates the ability to understand complex relationships between entities. Furthermore, users can set numeric and semantic similarity thresh- olds. Then, the system enables the user to "play" with various knobs and sliding bars to discover new insights and relationships between entities. As the user explore various knobs and changes various parameters, the system continuously notifies the user once the similarity between two entities fall below both the semantic and the numeric thresholds. SNSDS semi-automates the process of identifying relationships and saves the user a lot of time and effort to manually analyse large data sets.

As shown in Figure 9, the system gives the user a visual representation of the multidimensional numeric space with each entity represented as a point. The lines connecting two points represent the degree of numeric similarity be- tween the entities. The sliding bars (on the left side) gives the user the option to adjust the weights of various dimensions in the space to reflect the importance of this dimension in the similarity function. The graph clusters entities based on the numeric similarity between entities. The closer the entities on the graph, the more similar they are in the numeric space. The user sets (as shown in the figure) two thresholds, one for numeric similarity and another one for semantic similarity. As the user explores the domain by (a) adding/removing entities, entity groups and dimensions, (b) changing the weights of dimensions and (c) changing the numeric and semantic similarity thresholds, the system continuously alters the user of any pair of entities, where both their numeric and semantic similarities fall below the specified thresholds. Therefore, the tool helps the user explore the similarities across the numeric and semantic spaces at the same time.

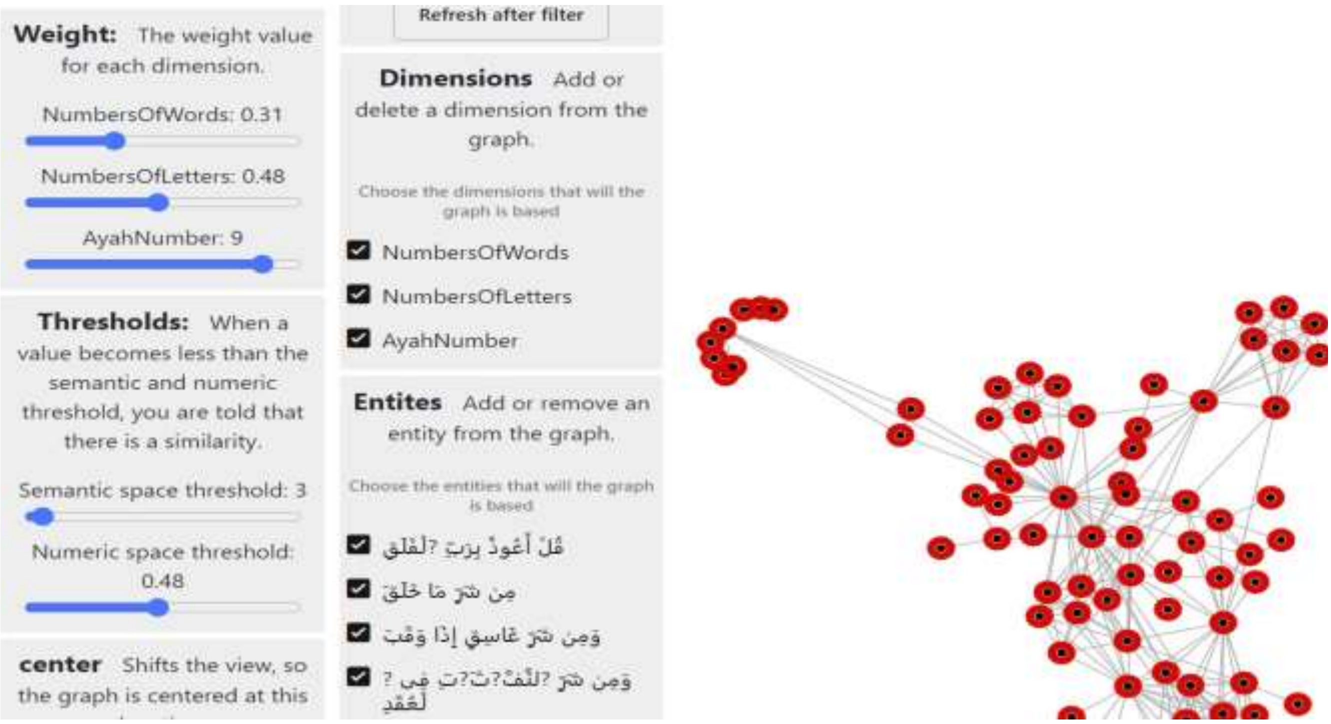

Fig. 9 Multidimensional Numeric Space Visualization with Entity Clusters and Relationship Dimensions

V. CONCLUSIONS

This paper proposed the SemSim, or Semantic Similarity, system to explore numeric and semantic similarities among heterogeneous entity groups. The sys- tem compares entities from different entity groups across different dimensions in a numeric space and correlates the numeric similarities with the semantic similarities that are given by a knowledge graph. As a case study, the proposed system has been used to explore the correlation between the semantic and numeric similarities for chapters, verses, and words of the text of the holy Quran. Overall, the proposed system, SemSim, bridges the gap between numeric and semantic spaces in multidimensional datasets. It also helps the user uncover and visualize correlated semantic and numeric patterns in large data sets.

ACKNOWLEDGEMENT

The authors hereby acknowledge the review support offered by the IJPCC reviewers who took their time to study the manuscript and find it acceptable for publishing

CONFLICT OF INTEREST

The authors declare that there is no conflict of Interest

REFERENCES

[1] R. Gisli Hjaltason and Hanan Samet. Index-driven similarity search in metric spaces (survey article). ACM Transactions on Database Systems (TODS), 28(4):517–580, 2003.

[2] N. Roussopoulos, Stephen Kelley, and Frederic Vincent. Nearest neighbor queries. In Proceedings of the 1995 ACM SIGMOD international conference on Management of data, pages 71–79, 1995.

[3] V. Gaede and Oliver Gu¨nther. Multidimensional access methods. ACM Computing Surveys (CSUR), 30(2):170–231, 1998.

[4] P. Rai and Shubha Singh. A survey of clustering techniques. International Journal of Computer Applications, 7(12):1–5, 2010.

[5] N. Vafaei, Rita A Ribeiro, and Luis M Camarinha-Matos. Normalization techniques for multi-criteria decision making: analytical hierarchy process case study. In doctoral conference on computing, electrical and industrial systems, pages 261–269. Springer, 2016.

[6] M. M. Alqahtani and Eric Atwell. Developing bilingual arabic- english ontologies of al-quran. In 2018 IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR), pages 96–101. IEEE, 2018.

[7] R. Ismail, Zainab Abu Bakar, and Nurazzah Abd Rahman. Extract- ing knowledge from english translated quran using nlp pattern. Jurnal Teknologi, 77(19), 2015.

[8] H. Kharrazi and Said Raghay. Collaborative ontology authoring in the domain of the holy quran knowledge. In 2019 7th Mediterranean Congress of Telecommunications (CMT), pages 1–3. IEEE, 2019.

[9] M. Alshammeri, Eric Atwell, and Mhd Ammar Alsalka. Quranic topic modelling using paragraph vectors. In Proceedings of SAI Intelligent Sys- tems Conference, pages 218–230. Springer, 2020.

[10] A.F. Huda, Moch R Deyana, QU Safitri, Wahyudin Darmalaksana, Ulfa Rahmani, et al. Analysis partition clustering and similarity measure on al-quran verses. In 2019 IEEE 5th International Conference on Wireless and Telematics (ICWT), pages 1–5. IEEE, 2019.

[11] E. Khadangi, Mohammad Moein Fazeli, and Amin Shahmohammadi. The study on quranic surahs' topic sameness using nlp techniques. In 2018 8th International Conference on Computer and Knowledge Engineer- ing (ICCKE), pages 298–302. IEEE, 2018.

[12] M. K¨oppen. The curse of dimensionality. In 5th online world conference on soft computing in industrial applications (WSC5), volume 1, pages 4–8, 2000.

[13] C. Sun Liew, Assad Abbas, Prem Prakash Jayaraman, Teh Ying Wah, Samee U Khan, et al. Big data reduction methods: a survey. Data Science and Engineering, 1(4):265–284, 2016.

[14] Y. Huang. Incorporating domain ontology information into clustering in heterogeneous networks. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 11(4):e1413, 2021.

[15] J. Liu, Zhe Wang, Jianmin Zhang, and Enhong Chen. A sur- vey of knowledge graph embedding: approaches and applications. In 2021 IEEE 23rd International Conference on High Performance Comput- ing and Communications; IEEE 19th International Conference on Smart City; IEEE 7th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), pages 1182–1187. IEEE, 2021.

[16] A. Salsabila Muhmad Rusli, Farida Ridzuan, Zulkifly Mohd Zaki, M Norazizi Sham Mohd Sayuti, and Rosalina Abdul Salam. A systematic review on semantic-based ontology for quranic knowledge. International Journal of Engineering and Technology (UAE), 2018.

[17] M. Nabeel Asim, Muhammad Wasim, Muhammad Usman Ghani Khan, Waqar Mahmood, and Hafiza Mahnoor Abbasi. A survey of ontology learning techniques and applications. Database, 2018, 2018.

[18] O. Romero and Alberto Abell´o. A survey of multidimensional modeling methodologies. International Journal of Data Warehousing and Mining (IJDWM), 5(2):1–23, 2009.