IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 18 July 2023, accepted 15 August 2023, date of publication 24 August 2023, date of current version 7 September 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3308514

RESEARCH ARTICLE

Emerging Exploration Strategies of Knowledge Graphs

MARWAN AL-TAWIL[1], VANIA DIMITROVA[2], DHAVALKUMAR THAKKER[3], AND BILAL ABU-SALIH[1]

[1]King Abdullah II School of Information Technology, The University of Jordan, Amman 11942, Jordan
[2]School of Computing, University of Leeds, LS2 9JT Leeds, U.K.
[3]Faculty of Science and Engineering, University of Hull, HU6 7RX Hull, U.K.

Corresponding author: Marwan Al-Tawil (m.altawil@ju.edu.jo)

ABSTRACT The utilization of semantic web technologies has led to the development of knowledge graphs represented as triples that allow for the exploration of specific and cross-domains. Despite the advantages of semantic links between entities in facilitating user exploration, they can also lead to an overwhelming number of exploration choices that can cause confusion, frustration, uncertainty, and a sense of being lost in the abundant graph, particularly for users who are not familiar with the domain. Thus, identifying exploration strategies is critical to improving user exploration and increasing exploration utility. This study aims to identify exploration strategies that promote knowledge utility (i.e., increase users' domain knowledge) and exploration experience (i.e., provide users with a positive and pleasant feeling). To accomplish this goal, an experimental user study was conducted, involving lay users in the musical instrument domain, where they were presented with an exploration task and then allowed to freely explore musical instruments. Parameters related to exploration paths were used to analyze the exploration patterns that users follow during their exploration. The findings reveal exploration strategies that promote knowledge utility and exploration experience. This research contributes to the literature on intelligent methods of guiding user exploration through knowledge graphs to enhance exploration effectiveness, which can have broad applications in knowledge graph utilization.

INDEX TERMS Knowledge graphs, knowledge utility, exploration experience, exploration paths, exploration strategies.

I. INTRODUCTION

Knowledge graphs (KGs) have evolved into fundamental elements in various exploration applications spanning from broad to specific domains [1]. A typical KG describes domain knowledge in the form of Resource Description Framework (RDF) triples [2]. Users are gradually being introduced to explore KGs, thereby enabling them to benefit from the vast knowledge contained within the graphs [3], [4], [5]. Unlike conventional search, where users have a clear idea of the expected results [6], exploratory search requires a significant amount of investigation and exploration effort (open-ended) [7] has unclear information needs [8], and is used for learning and exploration tasks [9]. Although semantic links in KGs support exploration through paths, they can also be a source of hindrance and confusion for users, especially for those who lack expertise in a particular domain. This can result in difficulties for users when making exploratory choices or formulating specific queries. Moreover, users may not be able to identify the most useful exploration paths for their needs or exploration tasks, ultimately leading to frustration, uncertainty, high cognitive load, confusion, and a sense of loss.

Initial efforts to facilitate KG exploration focused on providing users with textual or visual interfaces [10], [11]. Recent studies on KG exploration have united efforts from

The associate editor coordinating the review of this manuscript and approving it for publication was Mansoor Ahmed.

multiple fields such as human-computer interaction, user modeling and personalization, adaptive hypermedia, and the Semantic Web [1], [12], including personalized paths based on user preferences [13], a search-by-example paradigm by providing users with entities that are similar to a given search entity [14], domain summaries with data patterns [15], graph structure visualization [16], and constructing flexible queries to support user navigation [17]. Although a considerable amount of research has been conducted to enhance users' exploration experience of KGs, the focus has primarily been on investigative tasks and supporting learning through search [18], [19]. The objective of this learning perspective is to provide general tools to explore interlinked open educational resources [20].

Our focus is on supporting knowledge utility (i.e., expanding the user's domain knowledge while exploring a KG) and ensuring a positive exploration experience for the user. This is based on previous research, which acknowledges that when users explore unfamiliar domains, they unintentionally gain new knowledge about concepts or relationships that they were previously unaware of [21] and [22]. Our study in [23] was the first to develop an algorithm for generating exploration paths for knowledge expansion based on Ausubel's subsumption theory for meaningful learning [24]. This theory postulates that the human cognitive structure is hierarchically organized with respect to the levels of abstraction, where abstract and familiar concepts (i.e., knowledge anchors) are deliberately introduced to the user prior to the introduction of new concepts [29]. The new concepts then become anchored under relevant and abstract subsuming concepts, leading to meaningful learning [27].

The main goal of this study is to identify novel exploration strategies for enhancing users' domain knowledge and to provide users with a positive exploration experience. For this purpose, a task-driven experimental user study with lay users in the music domain was conducted, where users were presented with an exploration task and then asked to freely navigate through the KGs of a musical instrument to find information about a particular musical instrument with which they were unfamiliar. A key element for analyzing exploration paths followed by users is to identify path parameters for characterizing users' exploration and to examine the exploration patterns followed by users to identify exploration strategies. Path parameters included those related to the *level of generality in the class hierarchy*(i.e., top-level, middle-level, or bottom-level), parameters related to the *direction of exploration* (e.g., go-up, go-down, or staying at the same level in the class hierarchy), the average density of entities constructing an exploration path, *visiting knowledge anchors* (i.e., familiar concepts), and exploring dead-end entities.

The key contribution of this work is:

- Identify path parameters that affect the utility of the user's exploration.
- Analyze the effect of path parameters on knowledge expansion and exploration experience in a task-driven user study.
- Identify exploration strategies over KG to support knowledge utility and exploration experience.

The remainder of this paper is organized as follows. Section II discusses the related work. Section III provides the preliminaries and key definitions. Section IV describes the research methodology. Section V discusses the results and Section VI concludes the paper.

II. RELATED WORK

KGs have become increasingly popular because of their use as primary engines for various applications and their underlying abstract structures, which efficiently facilitate domain conceptualization and data management [27]. The variability of KGs is both a distinguishing feature and a barrier to their efficient implementation. Because of their extensive use, KGs are readily huge and complicated, and as a result, difficult to comprehend. Thus, even domain specialists find the content of KGs to be increasingly unfamiliar and nearly incomprehensible to novice users, necessitating the need for exploratory mechanisms to support the exploration over KGs. This section reviews the literature on KG exploration approaches. According to [1], KG exploration approaches can be divided into three main categories: profiling and summarization, exploratory analytics, and exploratory search.

A. PROFILING AND SUMMARIZATION

When lay users explore KGs, they may find it easier to experiment with different varieties, whereas users with more solid and consistent choices may be less inclined to stray from them [28]. In this context, KG profiling and summarization have been thoroughly investigated, and several methods and strategies have been suggested to demonstrate techniques for KG exploration in a clear and comprehensible manner [29]. For example, Sakurai et al. [30] developed a new method for artist recommendation that utilizes a KG to provide reasons for recommendations. By recommending artists instead of individual songs, multiple songs can be suggested simultaneously. A KG is constructed using listeners' listening histories and music track metadata, allowing for the representation of hierarchical relationships. Principe et al. [29] developed a minimalization-based profiling tool called ABSTAT-HD which can provide a profile for vast knowledge KGs. ABSTAT's modular design enables users to exploit the benefits of distributed computing. The authors characterized several datasets with varying degrees of complexity, including DBpedia and Microsoft Academic KG, to assess the scalability of the ABSTAT-HD. During the profiling process, three orthogonal factors were considered: the dataset size, its complexity in terms of the number of types, predicates, and ontological aspects, and the profiling type, which considers the total burden of the profiling process. The study in [31] offers personalized hint regression (PHR), a unique distillation technique that efficiently distills knowledge with diverse preferences. The PHR uses a personalization network to deliver tailored (or personalized) distillation for each user/item representation as a workaround for clustering. The customized network creates a mapping function that connects the instructor and student spaces, using the neighborhood information of each representation. Similarly, providing personalized KG exploration for education was reported in [32] and [33].

B. EXPLORATORY ANALYTICS

Attention has been paid to supporting sophisticated analytics for KGs for a while now [34]. Additionally, driving this is the growing need for both public and commercial institutions to represent business data in specific KGs [35]. In this context, Organisciak et al. [36] focused on the potential of exploratory data analysis and visualization tools to understand large bibliographic datasets. Hence, the HathiTrust+Bookworm was introduced, enabling a multi-faceted exploration of the HathiTrust Digital Library. This study situated this tool within the broader landscape of scholarly tools for exploratory data analysis. Although the construction of KGs has attracted great interest and has been applied to various domain applications, the investigation and visualization of the KG itself, which promotes interactive knowledge discovery and the creation of new theories, are not supported by computational tools or online frameworks. In this regard, the authors of [16] created a web framework called KG Exploration and Visualization (KGEV). This framework comprises of five steps: triple extraction, triple filtration, metadata preparation, knowledge integration, and graph database preparation. The system involves convenient user interface features that search a backend graph database, including node and edge search and filtering, data source filtering, neighborhood retrieval, and shortest-path computation. The proposed system enables the quick retrieval of pertinent texts that support the linkages in the KG, thus enabling human reviewers to assess the accuracy of the information collected. When examining vast and unfamiliar KGs, entity recommendations address the information overload problem. To address this issue, Yang et al. [37] proposed a KG exploration technique for topic-oriented entity recommendation. To determine the long-term negative intentions of a user, the authors constructed a negative-feedback memory network model. To achieve good intentions, the authors suggest a transformer-based sequence encoder. By combining positive and negative intentions using the proposed intent-attention technique, the system dynamically acquires adaptive intent. The incorporation of exploratory data analytics to draw conclusions from various datasets was also reported in [35].

C. EXPLORATORY SEARCH

Exploratory search refers to exploration that combines browsing and searching for knowledge acquisition [38]. This is also referred to as *exploration*, which requires a significant amount of investigation and exploration effort (open-ended) [38], has unclear information needs [39], and is used for learning and exploration tasks [40]. Faceted exploration is one of the first approaches that involves presenting a set of facets or attributes associated with entities in a KG, allowing users to select facets to narrow down their exploration [41], [42], [43]. In recent years, faceted exploration has been used as an approach for navigating through the KG, represented as RDF statements. For example, KTabulator was proposed in [44] as an interactive system that efficiently extracts, constructs, and expands ad hoc tables from extensive corpora over KGs. Users are allowed to extend their tables by choosing relevant entities and attributes. The authors of [45] proposed a relation facet approach for exploring KGs. Clustering methods are used to group relationships between entities based on entity similarity. PathWays was proposed in [46] as an interactive exploration tool for KGs to identify paths connecting specific entities in the KG to provide a comprehensive overview to the user. The closest work to this study is in [47], where the authors proposed a KG exploration method for lay users that depended on the graph structure. The proposed method utilizes a search tree algorithm to navigate the KG, where siblings are filtered to the most relevant nodes and ranked according to a specific criterion. However, our work suggests more comprehensive exploration strategies that consider several factors such as the direction of exploration, the level of abstraction in the class hierarchy, the density of entities, visiting highly inclusive and familiar entities (i.e., knowledge anchors), and exploring dead-end entities. Furthermore, the approach proposed in [47] was evaluated using a small and shallow class hierarchy, consisting of only 21 class entities, in comparison to our work which involved examining user exploration over two large class hierarchies: the Wind Instrument and the String Instrument (see Table 1).

Limitations of Existing Approaches: While current KG exploration approaches offer valuable insights for facilitating user exploration over KGs, they also come with certain limitations. For instance, profiling and summarization approaches heavily rely on diverse user preferences. If the user's preferences are not well-defined or change frequently, the distillation results might not accurately reflect the user's current interests. Existing exploration approaches assume that lay users are capable of making informed decisions about which facets, attributes, or entities to explore. However, in reality, lay users might not possess the necessary domain knowledge to select relevant facets or entities, potentially leading to suboptimal exploration outcomes, especially when lay users are confronted with numerous options and unfamiliar entities, leading to cognitive overload and confusion. Furthermore, current approaches lack exploration strategies that aid users to expand their understanding of the domain while ensuring a satisfactory level of exploration experience. In contrast, our work centers on providing exploration strategies tailored for uni-focal exploration, wherein the user initiates the exploration from a single entity within the KG and gradually traverses interconnected information. What distinguishes our approach is its explicit consideration of both the structure of the KG and the features associated with individual entities, which serves as a basis for suggesting exploration strategies.

Overall, the following are the key aspects that differentiate our approach from existing seminal works:

- Our approach specifically emphasizes enhancing users' domain knowledge and ensuring a positive exploration experience. This focus goes beyond traditional KG exploration, which often concentrates on data retrieval or analytics without considering the user's learning and satisfaction.
- We carry out a task-driven experimental user study in the music domain, involving lay users. This empirical approach allows us to gather real-world insights into how users interact with KGs and what strategies are effective in improving knowledge utility and exploration experience.
- Our approach identifies and analyzes a comprehensive set of path parameters related to exploration patterns, including generality in the class hierarchy, direction of exploration, density of entities, knowledge anchors, and dead-end entities. This thorough analysis provides a holistic view of user behavior during exploration.

III. PRELIMINARIES

We have provided the main definitions used in this study. Resource Description Framework (RDF) statements describe entities (i.e., nodes) and attributes (i.e., edges) in a KG, where each statement is a triple of the form <*Subject - Predicate - Object*> [43]. The *Subject* and *Predicate* denote the entities in the KG. An *Object* is either a Uniform Resource Identifier (URI) or string. Each *Predicate* URI denotes a directed relationship with *Subject* as a source entity and *Object* as a target entity.

Definition 1 (Knowledge Graph): A knowledge Graph $\boldsymbol{KG} = \langle \boldsymbol{V}, \boldsymbol{E}, \boldsymbol{T} \rangle$ is a labelled directed graph, depicting a set of RDF statements where $\boldsymbol{V} = \{v_1, v_2, \ldots, v_n\}$ is a finite set of entities; $\boldsymbol{E} = \{e_1, e_2, \ldots, e_m\}$ is a finite set of edge labels; $\boldsymbol{T} = \{t_1, t_2, \ldots, t_k\}$ is a finite set of RDF triples where each triple is a proposition in the form of $\langle v_u, e_i, v_o \rangle$ with $v_u, v_o \in V$, where v_u is the *Subject* (i.e., source entity) and v_o is the *Object*(i.e., target entity); and $e_i \in E$ is the *Predicate* (i.e., edge label).

In our analysis, the set of entities V mainly consists of the concepts (i.e., classes) of KG, and can also include individual objects (i.e., instances of concepts). Edge labels correspond to the semantic relationships between concepts and individual objects. These labels include the subsumption relationship `rdfs:subclassOf`, and `rdf:type` relationships. For a given entity v_i, we are primarily interested in its direct and inferred subclasses, and instances.

The *KG* has a *class hierarchy* consisting of all entities linked via the subsumption relationship `rdfs:subClassOf`. The set of entities V in the class hierarchy can be divided into the following three types by using the subsumption relationship `rdfs:subclassOf` (denoted as $\subseteq$) and following its transitivity inference into:

- *Root entity* (***r***) which is superclass for all entities in the class hierarchy.
- *Category entities* ($\boldsymbol{C} \subseteq \boldsymbol{V}$) which are the set of all inner entities (other than the root entity r) that have at least one subclass and may also include some individual objects.
- *Leaf entities* ($\boldsymbol{L} \subseteq \boldsymbol{V}$) which are the set of entities (other than the root entity r), that have no subclasses, and may have one or more individuals.

The set of edge labels E is divided further considering two relationship categories:

- *Hierarchical relationships*: the set of subsumption relationships between Subject and Object entities in corresponding triples.
- *Domain-specific relationships*: the relevant links in the domain, other than hierarchical links, for example, in the music domain, the instruments used in the same performance are related.

Definition 2 (Level of Generality): An entity $v \in C \cup L$ can be in one of three generality levels in the class hierarchy:

- *Top-level*: includes the set of category entities $C \subseteq V$ that are directly linked to root entity r via the subsumption relationship rdfs:subclassOf.
- *Middle-level*: includes the set of category entities $C \subseteq V$ that are not directly linked to the root r via the subsumption relationship rdfs:subclassOf.
- *Bottom-level:* includes set of leaf entities $L \subseteq V$.

Definition 3 (Entity Depth): The depth of an entity $v \in C \cup L$ is the length of the shortest path from entity v to root entity r in the class hierarchy via the subsumption relationship rdfs:subclassOf.

Definition 4 (Exploration Direction): The direction of an RDF statement $\langle v_i, e_i, v_{i+1} \rangle$, where e_i is the subsumption relationship rdfs:subclassOf, and can be one of the three directions:

- *go-up*: when the depth of Subject (v_i) > the depth of the Object (v_{i+1}).
- *go-down*: if the entity depth of Subject (v_i) < entity depth of Object (v_{i+1}).
- *stay*: When the depths of Subject (v_i) and Object (v_{i+1}) are equal, both belong to the same superclass.

Definition 5 (Exploration Path): An exploration path $\boldsymbol{P}$ in a $KG = \langle V, E, T \rangle$ is a sequence of finite set of RDF statements in the form of:

$P = \langle \langle v_1, e_1, v_2 \rangle, \langle v_2, e_2, v_3 \rangle, \ldots, \langle v_n, e_n, v_{n+1} \rangle \rangle$ where: $v_i \in V, i = 1, \ldots n+1$; $e_j \in E, j = 1, \ldots n$; v_1 and v_{n+1} are the first entity and the last entities of the exploration path P, respectively; n is the length of the exploration path P; $\langle v_i, e_i, v_{i+1} \rangle$ is an RDF statement (triple), where each exploration path P has n RDF statements.

In our analysis, the exploration path P is divided into three phases based on the triples constructing the path:

- *beginning of exploration:* characteristics related to Subject v_1, Predicate e_1 and Object v_2 of the first triple $\langle v_1, e_1, v_2 \rangle$ of an exploration path.

- *end of exploration:* characteristics related to the Subject v_n, Predicate e_n and Object v_{n+1} of the last triple $\langle v_n, e_n, v_{n+1}\rangle$ of an exploration path.
- *middle of exploration:* characteristics related to the Subject, Predicate and Object of the triples between the first triple $\langle v_1, e_1, v_2\rangle$ and the last triple $\langle v_n, e_n, v_{n+1}\rangle$.

IV. METHODOLOGY

A. APPROACH

Figure 1 shows the main steps of the proposed approach for identifying exploration strategies over KGs to promote knowledge utility and exploration experience.

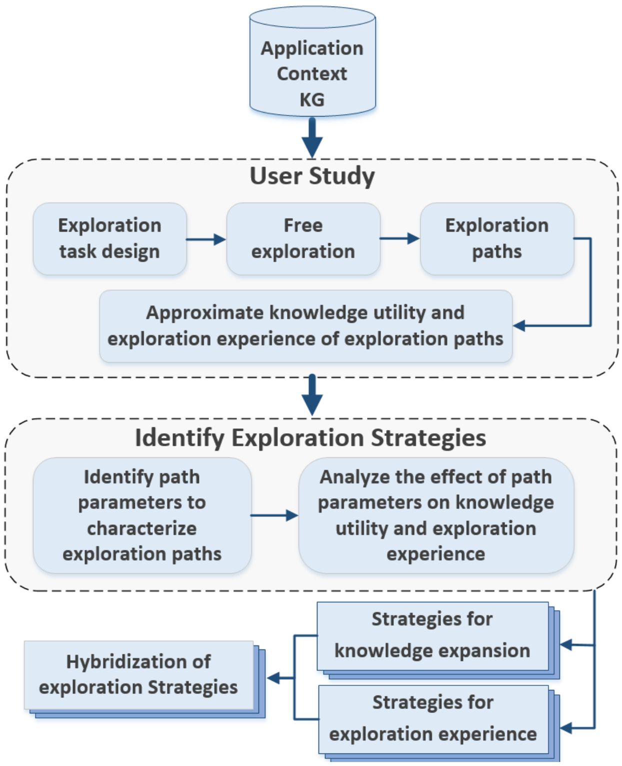

FIGURE 1. Structure of the proposed methodology.

The proposed approach comprises two main phases. In the first phase, a semantic data browser in a musical instrument domain called MusicPinta (described in subsection B) is utilized to validate user exploration through an experimental user study with lay users (described in subsection C). An exploration task is designed and presented to users. The exploration task allows participants to explore entities within an unfamiliar domain (e.g., Indian musical instruments), while placing them in a familiar situation (e.g., preparing a presentation about an Indian Wind Instrument called Bansuri). The participants are then asked to explore musical instruments freely using a semantic data browser. The knowledge utility and exploration experience values of the resultant exploration paths are approximated. In the second phase (described in section V), exploration path parameters, such as entity density and entity depth, are identified based on the structure of the KG. The effect of path parameters on knowledge utility cognitive processes (e.g., remember, categorize) is examined in order to identify exploration strategies to support knowledge expansion. Furthermore, The effect path parameters on exploration experience subjective processes (e.g., effort, frustration, performance) are examined to identify exploration strategies to enhance users' exploration experience. Finally, hybridization of exploration strategies is suggested.

B. APPLICATION CONTEXT

To validate user exploration, we utilize MusicPinta, a semantic data browser in the musical instrument domain [21]. Semantic data browsers are the first generation of semantic web applications that have emerged to support users while exploring KGs [48]. Similar to traditional web browsers (e.g., Google Chrome[1]), which allow users to follow hyperlinks while navigating through different webpages, semantic data browsers allow users to explore entities by following typed links expressed as RDF statements [49], [50], [51], [52]. They operate on semantically tagged content and present browsing links (edges) in the underpinning ontologies [21], [48], supporting the exploration of laypersons with uncertain information needs [53]. Semantic data browsers enable users to initiate an exploration session from a single entry point in a graph and move through different entities by following RDF links [21].

MusicPinta provides a unifocal interface for exploring information about musical instruments where exploration is restricted to a single start point (i.e., entity) and uses a resource at a time to navigate through the graph. At any time, the user focuses on one entity (the focus entity) from where links to other entities (candidate entities) can be chosen. At every juncture, when exploring a focus entity, the user must decide which candidate entities to choose for further navigation (i.e., to navigate through musical instrument information extracted from various linked datasets).

MusicPinta KG includes several sources, including DBpedia,[2] for musical instruments and artists extracted using SPARQL CONSTRUCT queries. The DBTune[3] dataset was used for music-related structured data. Among the datasets on DBTune.org we utilize: (i) Jamendo which is a large repository of Creative Commons licensed music; (ii) Megatune is an independent music label; and (iii) MusicBrainz is a community-maintained open source encyclopedia of music information. The dataset coming from DBTune.org (such as MusicBrainz, Jamendo and Megatunes) already contains the "sameAs" links between them for linking same entities. We utilize the "sameAs" links provided by DBpedia to link MusicBrainz and DBpedia datasets. Thus, DBpedia was linked to the remaining datasets from DBtune.org, enabling exploration via rich interconnected datasets. The MusicPinta

[1]https://www.google.com/chrome/index.html
[2]http://dbpedia.org/About
[3]http://dbtune.org/

dataset is available as an open source at the sourceforge.[4] All datasets in MusicPinta are available as RDF statements, and Music ontology[5] is the ontology used as the schema to interlink them. Overall, the dataset has 2.4M entities and 19M triple statements, taking 2GB of physical space, including 876 musical instrument entities, 71k (performances, albums, records, tracks), and 188k music artists. Table 1 lists main characteristics of MusicPinta dataset.

TABLE 1. Main characteristics of the musicpinta dataset.

Class hierarchy	Depth of class hierarchy	No. of classes	No. of DBpedia categories	No. of musical performance
String	7	151	255	348
Wind	7	108	161	1539
Percussion	5	82	182	127
Electronic	1	16	7	11
Other	1	7	0	2

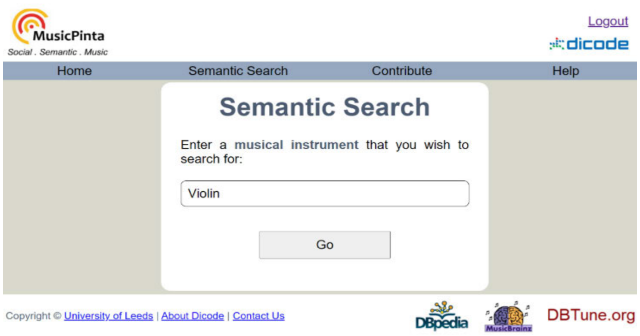

FIGURE 2. Semantic search interface in MusicPinta where a user inserts a name of a musical instrument (e.g., the entity Violin).

The MusicPinta KG includes five class hierarchies. Each class hierarchy has a number of classes linked via the subsumption relationship `rdfs:subClassOf` (e.g., 151 classes in the String Instrument class hierarchy). DBpedia categories are linked to classes in a hierarchy via dcterms:subject relationship, and classes are linked via the `domain-specific` relationship `MusicOntology:instrument` to musical performance. The depth of each class hierarchy is the maximum depth value for the entities in that class hierarchy. The underpinned music ontology provided sufficient class hierarchy for experimentation. For instance, the class hierarchies for String and Wind musical instruments have a depth of seven, which is considered ideal for analyzing user exploration patterns. The MusicPinta dataset provides an adequate setup for experimentation because it is fairly large and diverse, yet manageable. Figures 2–5 show examples of the user interface of MusicPinta.

[4]http://sourceforge.net/p/pinta/code/38/tree/
[5]http://musicontology.com/

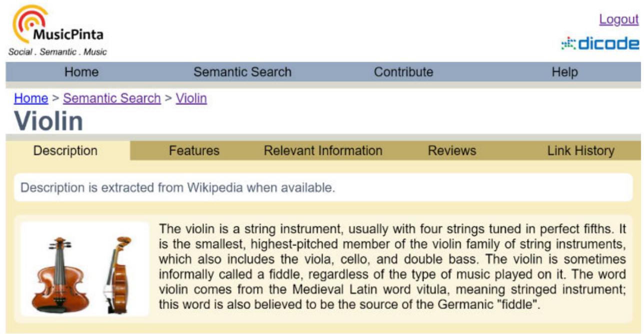

FIGURE 3. The description page of the entity 'Violin' in MusicPinta, extracted from DBpedia using CONSTRUCT queries.

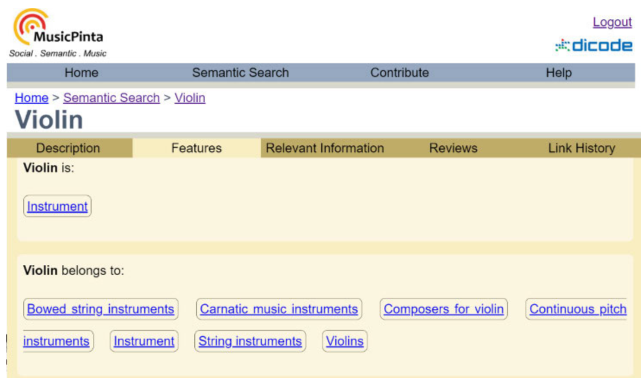

FIGURE 4. Semantic Links (i.e., predicates) related to the entity 'Violin' are presented in Features Information. It includes the semantic relationships rdf:type (e.g. Violin is an instrument), and semantic relationships rdfs:subclassOf (e.g., Violin is subclass of string instruments) and dcterms:subject to link an entity to its DBpedia category (e.g., Violin belongs to category Carnatic music instruments).

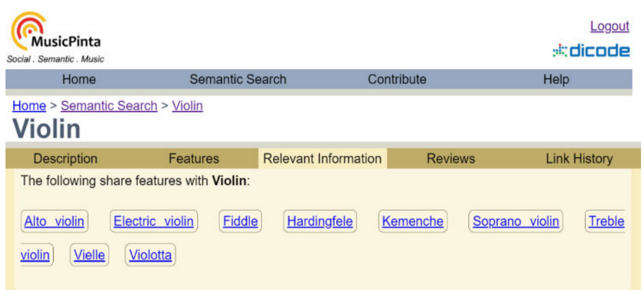

FIGURE 5. Semantic Links (i.e., predicates) related to the entity 'Violin' presented in Relevant Information, which include the semantic relationship rdfs:subclassOf (e.g. Violotta is a subClassOf Violin).

C. USER STUDY

We conducted a task-driven experimental user study with lay users (i.e., users who are not domain experts in the musical instrument domain) to identify the exploration patterns that support knowledge utility and exploration experience.

1) EXPLORATION TASK DESIGN

An important step in evaluating user exploration is the design of an appropriate exploration task [54]. According to [40], tasks that involve user exploration are classified as either investigative or learning-oriented, setting them apart from regular search-based tasks. Additionally, a typical exploration task must have certain characteristics. It should be broad in scope, not having any specific information needs in mind, and it should represent real-life tasks that occur in familiar situations. Also, the task should be discovery-oriented, venturing beyond users' current knowledge. Furthermore, the task must be open-ended, requiring a significant amount of exploration due to uncertainty or incomplete information. Lastly, the task should be set in an unfamiliar domain for the user [38], [54], [55]. A two-step approach (similar to the approach in [54]) was used to design the exploration task for study participants.

a: FIRST STEP

Design a task template that allows participants to explore entities within the unfamiliar topic, while placing them in a familiar situation. For instance, a scholar who intends to write an article about a new research topic. The task should be generic and encourage non-experts to seek knowledge in an unfamiliar domain. To achieve this, we propose designing a task template based on a general knowledge quiz show, allowing lay users to acquire as much information as possible. This will help to create a comfortable environment for participants to explore the topic and increase their knowledge. Inspired by the task templates in [54], we designed a task template to suit the musical instrument domain, as shown in Table 2.

TABLE 2. Task template used in the experimental user study.

Task template
"Imagine that you are a member of a team which will take part in a general knowledge quiz show. You have been asked to explore a musical instrument for 20 minutes to prepare a short presentation to describe to your team what you have learned about this instrument".

b: SECOND STEP

Identify unfamiliar candidate entities (e.g., find new research topics for a scholar) in a specific domain and incorporate the identified entities into the task template designed in step 1 above. For this, we conducted a questionnaire with lay users to identify unfamiliar entities in the `String Instrument` and `Wind Instrument` class hierarchies in the MusicPinta KG. These two class hierarchies were chosen because they have the richest class representation in terms of number of classes and class hierarchy depth (see Table 1). Furthermore, these two class hierarchies have the highest number of concepts (i.e., basic-level concepts) that are likely to be familiar to lay users [56]. There are nine basic concepts in the `String Instrument` class hierarchy (e.g., `Bouzouki, Guitar`, and `Violin`) and ten basic concepts in the `Wind Instrument` class hierarchy (e.g., `Accordion, Flute`, and `Recorder`) out of 24 anchors in the MusicPinta KG) [23]. We identified class entities belonging to the lowest quartile of the class hierarchy by selecting those at a depth of six or seven in two class hierarchies. The depth of both hierarchies was seven, as shown in Table 1. This idea is rooted in previous research in cognitive science that recognized the lack of familiarity of non-expert users with certain objects within a given domain [57]. A total of 61 class entities from the `String Instrument` and `Wind Instrument` class hierarchies were used in the questionnaire. We randomly assigned selected classes to 12 non-expert participants who had limited knowledge of musical instruments. These individuals may have encountered the instruments before, but none of them had any experience playing musical instruments. A 4-point Likert scale was used. Participants were asked to indicate their level of familiarity with the musical instruments by choosing one of the following options: (i) *High* (you have good knowledge and have played on the instrument); (ii) *Medium* (you have some knowledge and have listened to the instrument); (iii) *Low* (you have limited knowledge and have seen the instrument); (iv) *None* of the above. The questionnaire results showed that the most unfamiliar musical instruments from each class hierarchy were `Bansuri` (class hierarchy: `Wind Instrument`, origin: Indian) and `Biwa` (class hierarchy: `String Instrument`, origin: Japanese).

2) EXPERIMENTAL SETUP

a: PARTICIPANTS

32 participants were recruited voluntarily (compensation of £5 Amazon vouchers was offered to each participant. The number of participants in the current study is close to similar studies such as [3], [58], and [59]. The participants included 24 university students and eight professionals.[6] The participants' ages varied between 18 and 45 years (mean age is 30), and their cultural background (9 British, 3 Greek, 1 Austrian, 1 Chinese, 1 Italian, 5 Jordanian, 1 Libyan, 2 Malaysian, 6 Nigerian, 1 Polish, 1 Romanian and 1 Saudi). Furthermore, none of the participants was an expert in the music domain or had played musical instruments.

b: METHOD

We conducted four online surveys;[7] participants were randomly allocated to the online surveys; each survey had eight participants, and each participant was allocated one survey. Each participant freely explored one musical instrument (Biwa or Bansuri). The participants were directed to explore the same information during their exploration (i.e., participants explored 'Description,' 'Features' and 'Relevant information' of musical instruments – See figures 3 to 5). Each participant session was *conducted separately* and was observed by the authors. All participants were asked to provide feedback before, during, and after the interaction with the MusicPinta semantic data browser.

- *Task presentation [1 min]*: introduce the data exploration task to users at the outset of their exploration

[6] Academics and private Sector employees (Banking and Airlines).
[7] The study was conducted with Qualtrics (www.qualtrics.com).

sessions. For instance, the task template for exploring the Bansuri musical instrument could be: "Imagine that you are a member of a team that participates in a general knowledge quiz show. You have been asked to explore a musical instrument called Bansuri for 20 minutes to prepare a short presentation describing what you have learned about this instrument to your team.

- *Pre-study questionnaire [2 min]*: collect information about the participants' profiles and their familiarity with the music domain, focusing on the two musical instrument class hierarchies that will be explored: String Instrument and Wind Instrument. The participants' familiarity with the two class hierarchies varied from low to medium (63% and 78% for low familiarity with the String Instrument and Wind Instrument, respectively).
- *Free exploration [20 min]*: each user explores an unfamiliar instrument (Biwa or Bansuri). Each user was asked to explore five different entities. The number of entities explored is based on Miller's Law [60], which states that the number of objects that an average human can hold in working memory is 7 ± 2. We identified the knowledge utility and user exploration experience using user feedback on a modified version of the NASA-TLX questionnaire [61].

c: APPROXIMATING KNOWLEDGE UTILITY

Because schema activation [62] is used to assess a user's knowledge of a target concept X in the domain, the user is asked to name concepts that belong to and/or are similar to X. A schema activation tests conducted before and after exploration using three questions adapted from the well-known taxonomy by Bloom [63]. The Bloom taxonomy links human knowledge to six cognitive processes: remember, understand, apply, analyze, evaluate, and create. The cognitive processes: *remember* and *understand* are directly related to browsing and exploration activities. The process *remember* is retrieving relevant knowledge from long-term memory and includes recognition (locating knowledge) and recall (retrieving it from memory) [63]. Process *understand* concerns constructing meaning; the most relevant to our context is *categorize* (determining entity membership) and *compare* (detecting similarities) [63]. Accordingly, the following three questions were asked.

- *Remember*: What comes in your mind when you hear the word X?.
- *Categorize*: what categories does X belong to?.
- *Compare*: what concepts are similar to X?.

The number of *accurate* concepts named (e.g., naming an entity with its *exact name* or with a *parent* or *member* of the entity) by the user before and after exploration is counted, and the *difference indicates the knowledge utility of the exploration*. For example, if a user could correctly name *One* category (i.e., superclass), the musical instrument Bansuri belongs to (Q2) before an exploration, and then the user can correctly name *Three* categories to which the musical instrument Bansuri belongs after the exploration, then the effect of the exploration on the cognitive process *categorize* is indicated as 2 (i.e., as a result of the exploration, the user learned two new categories to which the musical instrument Bansuri belongs). If the user named only one instrument after the exploration, the user knowledge did not increase and the knowledge utility was counted as zero. Table 3 lists the statistics for three cognitive processes: remember, categorize, and compare.

TABLE 3. Statistics of knowledge utility cognitive processes in exploration paths in the task driven user study.

Cognitive process	Mean	Median	ST-DEV	Max	Min
Remember	2.13	2	1.01	4	0
Categorize	1.43	1	0.67	3	0
Compare	1.09	1	1.48	7	0

d: APPROXIMATING EXPLORATION EXPERIENCE

After each exploration, participants' feedback on the exploration experience, based on a modified version of the NASA-TLX questionnaire [60], was collected. The following six questions were presented to users to gather feedback on the user exploration experience, adapted from NASA-TLX:

- *Knowledge Expansion (KE)*: how much the exploration expanded your knowledge?
- *Content Diversity (CD)*: how diverse was the content you have explored?
- *Mental Demand (MD)*: how mentally demanding was this exploration?
- *Effort (EF)*: how hard did you have to work in this exploration?
- *Frustration (FR)*: how discouraged, irritated, stressed and annoyed you were in this exploration?
- *Performance (PE)*: How successful do you think you were in this exploration?

Table 4 lists values of the subjective processes adapted from NASA-TLX for each exploration path followed by the participants.

TABLE 4. Values of Subjective Processes adapted from NASA-TLX in the user study.

Subjective process	Mean	Median	ST-DEV	Max	Min
KE	4.72	5	2.02	9	1
CD	5.52	5	1.69	9	2
MD	3.28	3	2.29	8	0
EF	2.78	2	1.99	7	0
RF	1.66	1	2.19	9	0
PF	5.09	5	1.96	9	1

3) CHARACTERIZING EXPLORATION PATHS

We refer to the structure of a KG to identify the path parameters for characterizing exploration paths.

a: LEVEL OF GENERALITY
Class entities in a class hierarchy belong to three levels of generality: *top-level*, *middle-level* and *bottom-level* [64] (Definition 2 in Preliminaries). For example, figure 6 illustrates entities at the three generality levels on an extract from the `String Instrument` class hierarchy in the MusicPinta KG. For instance, the `String Instrument` (i.e., the root) and `Plucked String Instrument` (i.e., directly linked to rot) exist at the top-level of the class hierarchy. The class entity `Mandola` is a leaf entity at the bottom-level (i.e., Mandola has no subclasses). Finally, the class entities `Lute` and `Pipa` are both middle-level entities.

b: DIRECTION OF EXPLORATION
Depth of the source entity v_i(Subject entity) and the target entity v_{i+1}(Object) in a triple $\langle v_i, e_i, v_{i+1}\rangle$ of an exploration path P where e_i = `rdfs:subClassOf`, are used to determine the direction of exploration (Definition 4 in Preliminaries). Accordingly, three exploration directions were used to describe the direction of the triples constructing an exploration path: *go-up*, *go-down*, and s *tay*. For example, when a user in figure 6 travels from a subject entity `Plucked String Instrument` at *top-level* (depth = 1) to explore the object entity `Moon Lute` at the middle level (depth = 3), the direction of exploration is *go-down*.

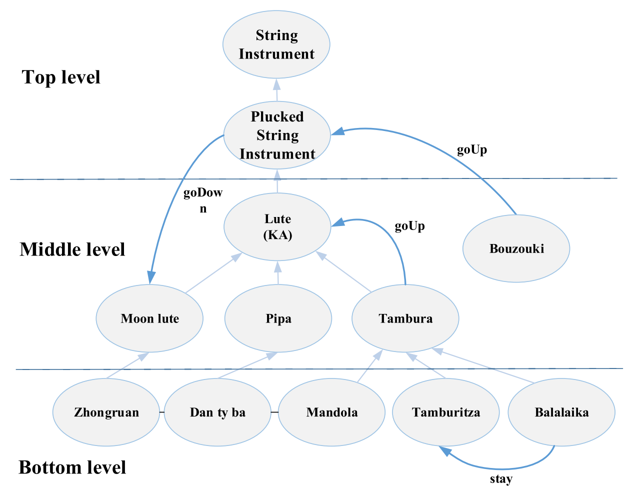

FIGURE 6. Extract from the String instrument class hierarchy in MusicPinta KG.

c: DENSITY
A scoping user study in [22] showed that exploring paths with dense entities is interesting and informative for users. We used the most common centrality algorithm, Degree Centrality, which considers importance based on the number of direct connections [65]. The subsumption relationship `rdfs:subClassOf` is used to measure the degree centrality of class entities in the class hierarchy. In particular, the degree centrality of a node is the sum of its subclasses and super-classes. The average density of entities in an exploration path can be categorized as high-density (≥ mean average) or low-density (< mean average).

d: KNOWLEDGE ANCHORS
Knowledge anchors (KA) represent highly inclusive and familiar entities at the basic level of abstraction in a human cognitive structure, from where links to new entities can be made to facilitate knowledge expansion. Our work in [23] has developed algorithms for identifying knowledge anchors in the MusicPinta KG.

e: DEAD-END ENTITIES
The construction of the MusicPinta KG included DCT subjects that had no connections to other entities in the KG. Users who explored dead-end entities had no choice but to proceed with their exploration and returned to the previous entity to choose another entity to explore.

V. RESULTS AND DISCUSSION
In this section, we describe how the exploration paths followed by the participants in the user study were analyzed to identify two groups of exploration strategies: (i) exploration strategies aimed at expanding users' domain knowledge, and (ii) exploration strategies aimed at providing the user with a positive exploration experience.

TABLE 5. Statistically significant differences of cognitive processes (Mann-Whitney, 1-tail) of two groups of exploration paths based on the mean values of path parameters in Table 3: exploration paths with path parameter values less than mean value; AND EXPLORATION paths with path parameter values greater than or equal the mean).

Path parameter	Remember P-value	Categorize P-value	Compare P-value
top-level	<0.01(Pos)	<0.0001(Pos)	0.20
middle-level	0.02(Pos)	0.30	0.24
bottom-level	0.01(Neg)	0.053	0.37
go-up	0.12	0.43	0.49
go-down	0.08	0.045(Pos)	0.44
stay	0.45	0.14	0.09
density	0.01(Pos)	<0.0001(Pos)	0.13
KA	0.06	0.13	0.48
dead-end	0.24	0.02 (Neg)	0.23

A. EXPLORATION STRATEGIES FOR KNOWLEDGE EXPANSION
To identify exploration strategies for knowledge expansion, we are examining the impact of path parameters on three cognitive processes of knowledge utility: remember, categorize, and compare. Analyzing the effects of path parameters with a positive impact will aid in identifying exploration strategies likely to enhance the user's domain knowledge. Conversely, investigating the effects of path parameters that negatively impact cognitive processes will help pinpoint exploration strategies that are likely to minimize the acquisition of new knowledge. Table 5 illustrates the effects of path parameters on the three cognitive processes of knowledge utility.

1) EXPLORATION STRATEGIES TO SUPPORT THE REMEMBER COGNITIVE PROCESS

a: PATH PARAMETERS WITH A POSITIVE EFFECT ON THE REMEMBER COGNITIVE PROCESS

The results in Table 5 indicate a positive effect on the "remember" cognitive process for three path parameters: top-level, middle-level, and density. To further examine the effect of path parameters on the remember cognitive process, we observed the patterns of path parameters on two groups of exploration paths: (i) exploration paths with high remember values ($\geq$ the mean value in Table 3) to identify exploration strategies to support the remember cognitive process; and (ii) exploration paths with low remember values (i.e., $<$ the mean value in Table 3) to identify exploration strategies with negative impact on the remember cognitive process (i.e., exploration strategies that have to be avoided).

b: EFFECT OF EXPLORING TOP-LEVEL ENTITIES

The exploration patterns presented in figure 7 display the ratios of top-level entities to entities constructing exploration paths with high and low remember values. The figure shows that users did not explore top-level entities in v1 (i.e., users started their exploration from bottom-level entities in the class hierarchy) nor v3 and v6 (i.e., users explored entities at the bottom-level or middle-level in the class hierarchy). The findings indicated that exploring top-level entities at the start and end of an exploration path facilitates remembering new entities. Conversely, exploring top-level entities in the middle of exploration, such as the high ratios of top-level entities in entities v4 and v5, had a lower remembering utility.

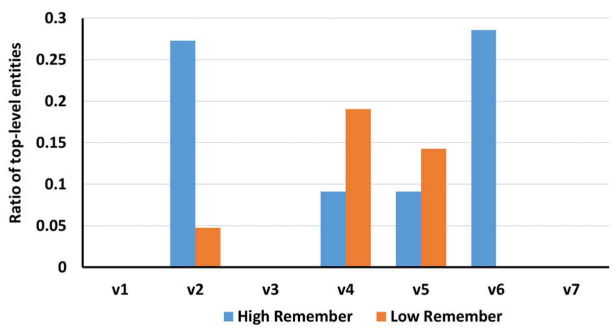

FIGURE 7. Ratio of top-level entities of exploration paths with high and low remember values.

Spearman's correlation analysis indicated a strong positive correlation between top-level entities and the average density (r = 0.81, p = 0). Consequently, we observed patterns in the average density of top-level entities, as shown in figure 8. The outcomes revealed that users could recall more knowledge when they explored top-level entities with a high density, especially at the start of the exploration path. For instance, when lay users explore a new subject area, they may encounter many options in the KG. To make sense of this, users may begin by exploring the most general, top-level entities (such as overarching topics or themes) at the beginning

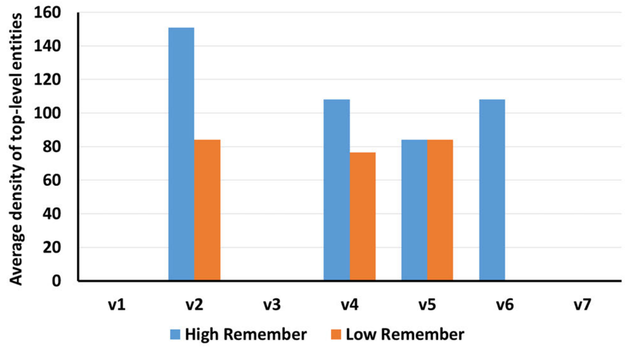

FIGURE 8. Average density of top-level entities for paths with high and low remember values.

of the exploration process. This provides a high-level understanding of the subject area before delving into more specific details. However, if lay users begin exploring top-level entities in the middle of the process (such as diving into specific topics before grasping the bigger picture), they may struggle to retain new knowledge. The average density in v1, v3, and v7 was zero because the participants did not visit the top-level entities (see figure 7).

c: EFFECT OF EXPLORING MIDDLE-LEVEL ENTITIES

According to figure 9, users who began their exploration by visiting entities at the middle-level in the class hierarchy had lower remember values than those who started with top-level entities (as shown in figure 7). However, the outcomes presented in figure 9 indicate that visiting middle-level entities in the middle of the exploration period has a beneficial impact on remembering new knowledge. This suggests that starting with a higher level and gradually drilling to lower levels may be a useful strategy for remembering new knowledge. That is, if users started at a top-level entity and drilled into a middle-level entity, they are more likely to remember information about the middle-level entity than if they started at the middle-level entity directly.

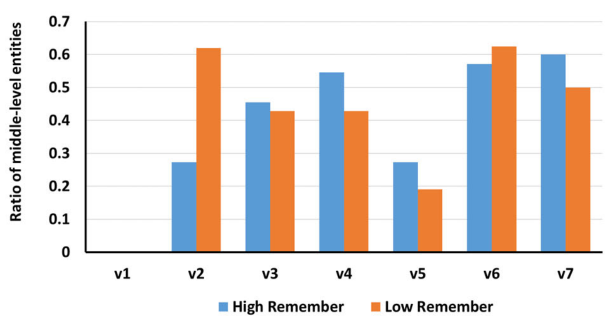

FIGURE 9. Ratio of middle-level entities for paths with high and low remember values.

d: EFFECT OF DENSITY

Figure 10 illustrates that initiating exploration with dense entities enhances the retention of remembering new

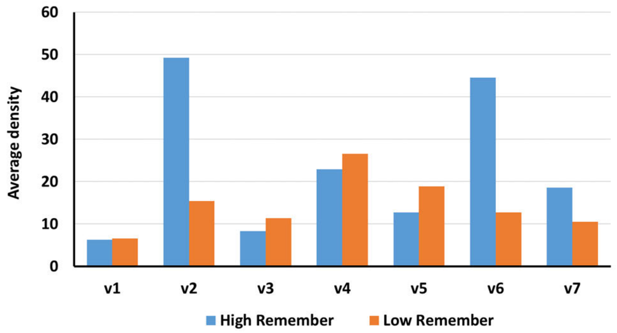

FIGURE 10. **Average density of exploration path entities with high and low remember values.**

knowledge, whereas exploring dense entities during the middle stage of exploration hinders the cognitive process of remembering. Exploring dense entities at the beginning of exploration can support remembering new knowledge because it allows the learner to build a foundation of knowledge that makes it easier to understand subconcept concepts later. By visiting less dense concepts, the learner can then develop a basic understanding of the subject matter and establish connections between new information and prior knowledge. This can help to create a cognitive structure that makes it easier to remember and retain new information. Furthermore, these findings are compatible with previous observations (figures 7–9), which indicate that users who explore high-density top-level entities recall more concepts than those who start with middle-level entities. Overall, these strategies suggest that learners should focus on building a foundation of knowledge by exploring dense entities at the beginning of exploration, then move on to less dense concepts and avoid dense entities during middle of exploration.

e: PATH PARAMETERS WITH A NEGATIVE EFFECT ON THE REMEMBER COGNITIVE PROCESS

According to the results in Table 5, exploring *bottom-level entities* negatively affects the remember cognitive process. Exploration patterns in figure 11 show that exploring bottom-level entities (i.e., leaf entities with no subclasses) has a detrimental impact on the remember cognitive process. The negative effect on the remember cognitive process associated with exploring bottom-level entities can be explained by the fact that our brains tend to remember information better when presented in a structured, hierarchical manner. When we explore bottom-level entities without first understanding the broader context or abstract concepts, we are likely to forget the information we learned, as they are not anchored to familiar and abstract concepts at a higher level, because our brains tend to remember information better when it is presented in a structured, hierarchical way. Furthermore, the patterns in figure 11 shows that all users went from bottom-level in v1 to explore middle-level or top-level entities in the class hierarchy in v2. This shows that lay users prefer to explore more generic terms at the beginning of their exploration rather than staying at the bottom-level.

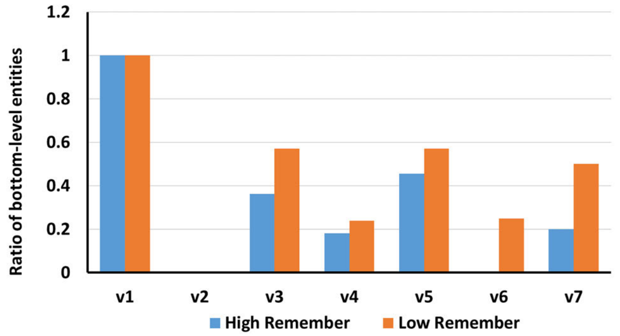

FIGURE 11. **Ratio of bottom-level entities for paths with high and low remember values.**

f: EXPLORATION STRATEGIES TO SUPPORT REMEMBERING NEW KNOWLEDGE

Based on the exploration patterns mentioned above, the following exploration strategies can be inferred for path parameters:

Exploring top-level entities:

- Start the exploration path by exploring top-level entities. When exploring a new subject area, it is recommended to start an exploration path by exploring the most general top-level entities at the beginning of the exploration process. This provides a high-level understanding of the subject area before delving into more specific details.
- Explore the top-level entities at the end of the exploration path. Hence, users can see more useful entities and links that help them establish meaningful links that can aid in the remember cognitive process.
- Avoid exploring top-level entities in the middle of the exploration path. Exploring top-level entities in the middle of exploration, such as diving into specific topics before getting a grasp of the bigger picture, may lower remembering utility. Therefore, it is recommended to avoid exploring top-level entities in the middle of the process and instead focus on specific topics after establishing a high-level understanding of the subject area.

Exploring middle-level entities:

- It is recommended that middle-level entities be explored in the middle of the exploration path. However, exploring middle-level entities at the beginning of exploration is not recommended. According to figure 9, starting at the top-level and gradually drilling to the lower levels may be a useful strategy for remembering new knowledge. Therefore, it is recommended to start with top-level entities and then drill down to middle-level entities before exploring the specific topics.

Exploring bottom-level entities:

- Avoid exploring bottom-level entities through the exploration path, particularly laypersons. However, if users (especially domain experts) want to explore bottom-level entities, they should first grasp a good understanding of the domain by exploring dense entities at a more abstract level (e.g., users at *v*2 in figure 11 do not explore the bottom level). In acc cases, the users avoid the bottom level at the end of the exploration path.

Average density:

- Initiate exploration with high-density entities at the top-level of the class hierarchy. This can enhance the retention of new information because it helps build a foundation of knowledge that makes it easier to understand sub-concepts later, especially for learners who are new to the domain and want to establish a basic understanding before delving deeper into a class hierarchy.
- Explore less-dense concepts after building a foundation. Once the learner has built a foundation for knowledge by exploring dense entities, they can move on to less dense concepts. This can help to establish connections between new information and prior knowledge, which can aid in creating a cognitive structure that makes it easier to remember and retain new information.
- Explore dense entities at the end of the exploration so that users can have a better understanding of the domain and remember more things by creating useful links between entities.

The overall characteristics of the exploration strategies to support the remember cognitive process are summarized in Table 6.

TABLE 6. summary of characteristics of exploration strategies to support the remember cognitive process.

Path parameter	Beginning of exploration path	Middle of exploration path	End of exploration path
top-level	√	-	√
middle-level	-	√	-
bottom-level	-	-	-
go-up	√	-	-
go-down	-	√	-
stay	-	-	-
high density	√	-	√
low density		√	-
KA	-	-	-
dead-end	-		

2) EXPLORATION STRATEGIES TO SUPPORT THE CATEGORIZE COGNITIVE PROCESS

a: PATH PARAMETERS WITH A POSITIVE EFFECT ON THE CATEGORIZE COGNITIVE PROCESS

The results in Table 5 show that exploring top-level entities and density has a significant positive impact on the cognitive process of categorization. Additionally, this study revealed that moving down the class hierarchy aids the categorization of new knowledge in human cognitive structures. To determine the exploration strategies that would facilitate the categorization process, this study analyzed the patterns of path parameters in exploration paths with high categorization values (to identify favorable strategies) and low categorization values (to identify strategies to be avoided).

Effect of Exploring Top-Level Entities As depicted in figure 12, exploring top-level entities via exploration paths assists in categorizing new knowledge at higher rates compared to exploring entities at other levels of generality in the class hierarchy. The figure shows that exploring top-level entities will always support categorizing new entities in human cognitive structures, especially because top-level entities are abstract and familiar concepts (e.g., the concept of a String Instrument) with many subclasses (e.g., a String Instrument has 150 subclasses). Hence, users can categorize concepts (e.g., the Guitar belongs to the String Instrument category). Nevertheless, the study findings revealed that users did not engage in the exploration of top-level entities in *v3*. This indicates that individuals who explored top-level entities in *v2* proceeded to explore entities at the middle-level or bottom-level of the class hierarchy.

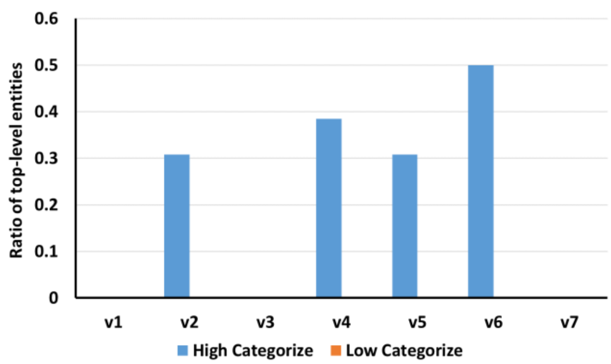

FIGURE 12. Ratio of top-level entities for paths with high and low categorize values.

Correlation analysis indicated a strong positive relationship between the average density and top-level entities, with a Spearman correlation coefficient of 0.81 and a p-value of 0. Consequently, the average density of top-level entities in figure 13 was examined in exploration paths with both high and low categorized values. The findings revealed that the top-level entities visited by users at the start of the exploration had higher density values than those at other points in the exploration path. For instance, in *v2*, four users explored top-level entities, with three examining `String Instrument` (the root entity with a density of 150), and one user explored `Plucked String Instrument` (directly linked to the root entity with a density of 84). Conversely, the density of top-level entities in the middle of exploration was lower than that at the start of exploration. For example, in v4, five users explored top-level entities, with one examining `Wind Instrument` (a root entity with a density of 108), three exploring `Woodwind` (directly linked to the root `Wind`

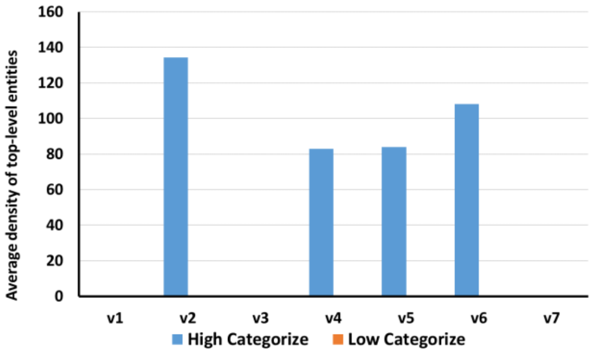

FIGURE 13. Average density of top-level entities for paths with high and low categorize values.

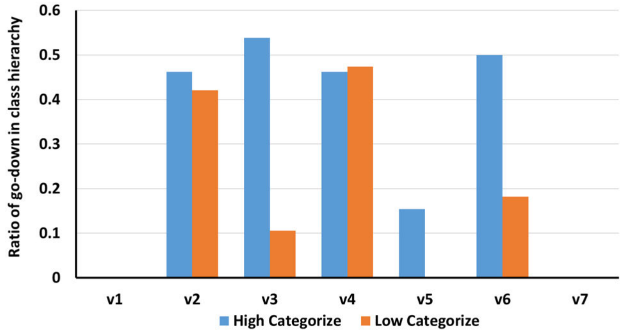

FIGURE 15. Ratio of going down in exploration path entities with high and low categorize values.

`Instrument` entity with a density of 74), and one exploring `Plucked String Instrument`.

Effect of Density: Figure 14 shows the impact of average density on the cognitive process of categorization. The findings indicate that dense entities facilitate their categorization, particularly at the start and end of an exploration path. When entities with high density are explored, users can establish connections with many subclass entities at lower levels in the class hierarchy. Furthermore, the exploration pattern in figure 14 shows that with an increase in the average density of entities along the exploration path, users can categorize higher rates of concepts in their cognitive structures.

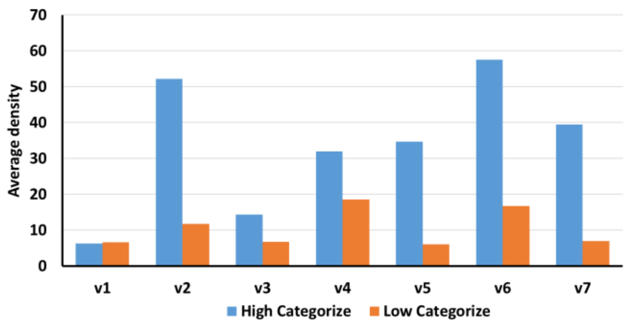

FIGURE 14. Average density of exploration path entities with high and low categorize values.

Effect of Go-Down: The findings presented in figure 15 indicate that users who descended the class hierarchy were better able to categorize new knowledge than those who did not. This effect was found to be statistically significant at v2, which refers to the point at which users transitioned from exploring top-level entities (i.e., broad categories) to examining middle-level entities (i.e., subcategories). Going down in the class hierarchy can be a useful strategy for exploring KGs because it allows users to focus their attention on more specific concepts and relationships. By exploring middle-level entities, users can gain a deeper understanding of the underlying structure and organization of the KG. This can help them categorize and make sense of new information more effectively. This observation was significant at the entity level when users went down from entity *v2* (i.e., users explored high-density entities (see figure 13) to entity *v3* (i.e., users went down to explore entities below the top level – no top-level entities at *v3* in figure 13).

b: PATH PARAMETERS WITH A NEGATIVE EFFECT ON THE CATEGORIZE COGNITIVE PROCESS

According to the results in Table 5, exploring dead-end entities has been observed to negatively affect the cognitive process of categorization. The results presented in figure 16 suggest that exploring dead-end entities negatively affects the categorization of concepts within human cognitive structures. Dead-end entities are those that do not have any connections to other entities within the KG; therefore, exploring them can be less productive in expanding domain knowledge of users. Exploring dead-end entities may distract users from identifying and categorizing the relevant entities within a KG. Users who have explored higher rates of dead-end entities had lower categorized values, especially towards the end of the exploration path (e.g., users had low categorization values when they explored dead-end entities in vertices v4 to v7). Nevertheless, the results in figure 16 show that some participants who explored dead-end entities at the beginning of the exploration were able to categorize entities, especially at v2. Upon examining the characteristics of v2, we noticed

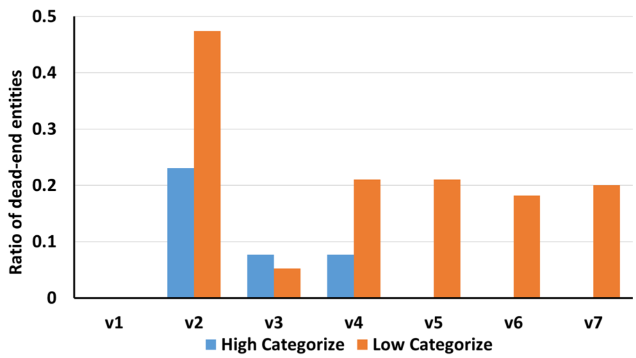

FIGURE 16. Ratio of exploring dead-end entities in exploration path entities with high and low categorize values.

that the average density of entities in v2 was relatively high (see figure 14). Hence, although the users explored high rates of dead-end entities in v2, exploring a few highly dense entities is sufficient to support the categorization cognitive process.

c: EXPLORATION STRATEGIES TO SUPPORT CATEGORIZING NEW KNOWLEDGE

The above exploration patterns suggest the following exploration strategies based on the path parameters:

Exploring top-level entities:

- Explore the top-level entities at the beginning of the exploration path. The findings indicated that the top-level entities visited by users at the start of exploration had higher density values than those at other points in the path.
- Consider the density values of the top-level entities. Correlation analysis indicated a strong positive relationship between average density and top-level entities. This will help users to identify important entities and their relationships.

Average density of entities:

- Start and end exploration paths with dense entities. The findings indicate that dense entities facilitate their categorization, particularly at the start and end of an exploration path. Users can begin exploring the KG by selecting a dense entity relevant to their area of interest. Similarly, they can end their exploration by focusing on dense entities to see more useful entities and links and to establish meaningful links that can air the categorization process.
- Explore high-density entities through an exploration path. Exploring high-density entities can help users establish connections with a large number of subclass entities at lower levels in the class hierarchy.
- Increase the average density of entities along the exploration path. Increasing the average density of entities along the exploration path helps users to categorize higher rates of concepts.

Go-down:

- Go-down from top-level entities to middle-level and bottom-level entities. The findings revealed that individuals who explored top-level entities subsequently proceeded to explore entities at the middle-level or bottom level of the class hierarchy.

Dead ends:

- Avoid dead-end entities: The results suggest that exploring dead-end entities can negatively affect the categorization cognitive process within human cognitive structures. Therefore, users should avoid exploring dead-end entities as much as possible.

The overall characteristics of the exploration strategies used to support the categorization of concepts in cognitive processes are summarized in Table 7.

TABLE 7. summary of characteristics of exploration strategies to support the categorize cognitive process.

Path parameter	Beginning of exploration path	Middle of exploration path	End of exploration path
top-level	√	-	√
middle-level	-	√	-
bottom-level	-	-	-
go-up	√	-	-
go-down	-	√	-
stay	-	-	-
high density	√	√	√
low density		√	-
KA	-	-	-
dead-end	-		

3) EXPLORATION STRATEGIES TO SUPPORT KNOWLEDGE EXPANSION

In this section, we identify exploration strategies that support remembering and categorizing concepts for knowledge expansion.

Levels of generality in class hierarchy.

- Explore top-level entities at the beginning of the exploration path to provide a high-level understanding of the subject area before delving into more specific details.
- Explore middle-level entities in the middle of the exploration path.
- Explore top-level entities at the end of the exploration path to establish meaningful links that can aid in the remember cognitive process.
- Consider the density values of the top-level entities to identify important entities and their relationships, particularly at the beginning and end of the exploration path.
- Avoid exploring bottom-level entities through the exploration path, especially laypersons.

Direction of exploration

- Go-down from top-level entities to middle-level entities in the middle of exploration. Then, we go-up towards the top-level entities.

Density of entities.

- Initiate exploration with high-density entities at the top-level of the class hierarchy to enhance the retention of new information and build a foundation for knowledge hierarchy.
- Explore entities with high or low densities in the middle of the exploration path. If low-density entities are explored, the average density of entities along the exploration path increases.
- Explore dense entities at the end of the exploration to gain a better understanding of the domain and remember more things by creating useful links between entities.

Dead ends:

- Avoid dead-end entities along the exploration path.

B. EXPLORATION STRATEGIES TO SUPPORT EXPLORATION EXPERIENCE

To identify exploration strategies for supporting the exploration experience, we examine the effect of path parameters on the six subjective processes derived from Nasa-TLX: knowledge expansion (KE), content diversity (CD), mental demand (MD), effort (EF), frustration (FR) and performance (PE). Examining the effect of path parameters with a positive impact will help identify exploration strategies that are likely to provide users with pleasant and positive experience. Whereas examining the effect of path parameters with a negative impact will help identify exploration strategies that are likely to minimize the exploration experience. Table 8 illustrates the effect of the path parameters on the six subjective processes.

TABLE 8. Statistically significant differences of subjective processes (Mann-Whitney, 1-tail) of two groups of exploration paths based on the mean values of path parameters in Table 3: exploration paths with path parameter values less than mean value; and exploration paths with path parameter values greater than or equal the mean).

Path parameter	KE P value	CD P value	MD P value	EF P value	FR P value	PE P value
top-level	0.25	0.18	0.12	0.46	0.46	0.11
middle-level	0.49	0.27	0.24	0.06	0.31	0.37
bottom-level	0.11	0.01(Neg)	0.02(Neg)	0.12	0.46	0.28
go-up	0.15	0.36	0.34	0.17	0.09	0.01(Pos)
go-down	0.42	0.47	0.22	0.34	0.38	0.19
stay	0.33	0.13	0.13	0.33	0.39	0.11
density	0.37	0.06	0.08	0.39	0.24	0.42
KA	0.39	0.03(Pos)	0.12	0.36	0.22	0.35
dead-end	0.42	0.17	0.14	0.30	0.39	0.13

The results in Table 8 show that the three subjective processes were influenced by three parameters related to the exploration paths. Specifically, KA has a favorable impact on CD, whereas exploring bottom-level entities has an unfavorable impact on CD. The exploration patterns depicted in figure 17 illustrate that increasing KA has a positive effect on CD throughout the exploration path. In other words, as users encounter more instances of KA during exploration, they tend to experience higher levels of content diversity. Moreover, the effect of KA on CD became more pronounced when users were exposed to higher rates of KA. This means that the positive impact of KA on CD is not constant; rather, it increases as users encounter more instances of KA during exploration. Furthermore, the patterns displayed in figure 18 demonstrate that exploring bottom-level entities has a detrimental effect on CD. As users explore more of these entities, they tend to experience lower CD levels.

Mental demand (MD): The patterns presented in figure 19 demonstrate that when users visited leaf entities located at the bottom of the class hierarchy, they required less effort to perform the exploration task compared to exploring other entities located in the middle or top levels of the class hierarchy. This suggests that exploring lower levels of a class hierarchy can be more efficient and less time-consuming for users, particularly when they must quickly acquire specific knowledge related to the domain. The results emphasize the

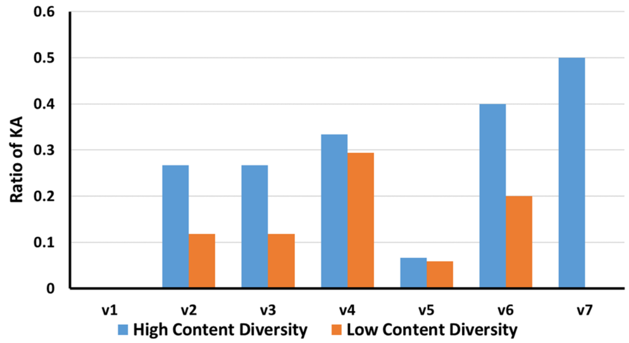

FIGURE 17. Ratio of exploring knowledge anchors through paths with high and low contend diversity values.

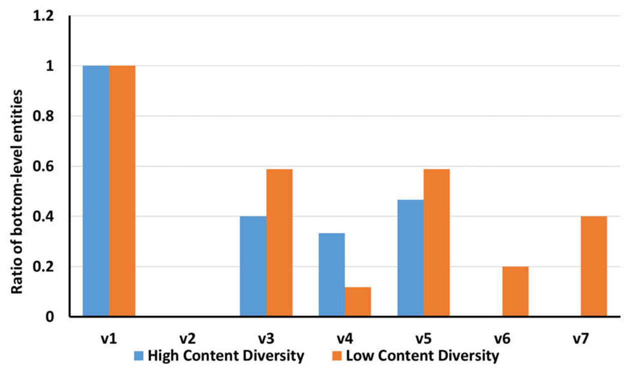

FIGURE 18. Ratio of exploring bottom-level entities through paths with high and low contend diversity values.

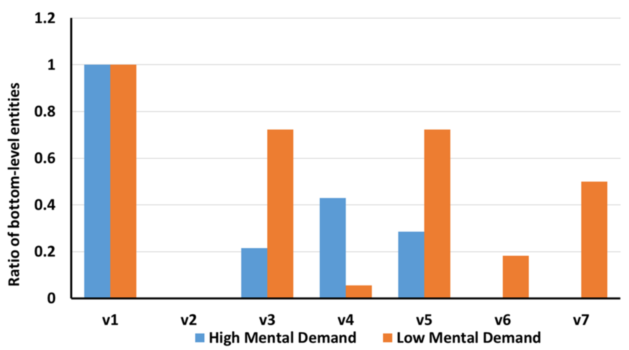

FIGURE 19. Ratio of exploring bottom-level entities through paths with high and low mental demand values.

practical implications of the findings, particularly for domain experts. Domain experts often need to explore class hierarchies to gain a deeper understanding of the domain and to acquire new knowledge. Knowing that the bottom level of the class hierarchy requires less effort to explore, domain experts can focus their exploration efforts on these areas to obtain the information they need quickly. This can save time and improve the efficiency of the exploration process, ultimately leading to better decision-making and problem-solving in the domain. However, individuals are unfamiliar with class hierarchy structures. Starting at the bottom level provides a more accessible entry point into the hierarchy and makes it easier for individuals to navigate and explore different entities. This can be particularly beneficial for individuals unfamiliar with domain-specific concepts in the class hierarchy.

Performance (PE): The patterns displayed in figure 20 indicate that users initially increased their going-up pattern during the early stages of exploration but decreased it during the middle phase before increasing again towards the end. Therefore, an effective exploration strategy to enhance task performance would be to begin the exploration by ascending to the top-level and subsequently alternate between upward and downward movements towards the end of the exploration.

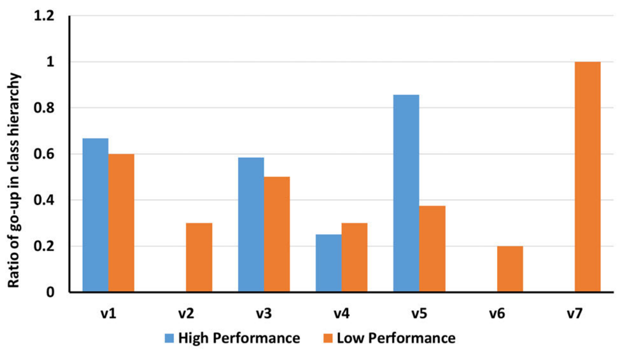

FIGURE 20. Ratio of going up through paths with high and low performance values.

a: EXPLORATION STRATEGIES TO SUPPORT EXPLORATION EXPERIENCE

The above exploration patterns suggest the following exploration strategies based on the path parameters:

Levels of generality in class hierarchy

- Start the exploration path by exploring entities in the bottom-level of the class hierarchy to decrease mental demand (MD). The results demonstrated that exploring the lower levels of the class hierarchy requires less effort than exploring the other levels. Furthermore, starting from the bottom-level can provide a more accessible entry point into the class hierarchy for lay users, and make it easier for them to navigate and explore different entities. However, for domain experts, focusing on their exploration efforts at the bottom-level can be an efficient way to quickly obtain the information they need.
- Focus on exploring entities at the top and middle levels in the class hierarchy to increase CD. By exploring higher-level entities, users can gain a broader understanding of the subject matter they are exploring, which can help them to identify new areas of interest and diverse content.

Direction of exploration:

- Start by sending users to the top-level of the class hierarchy to support the task performance. Sending users toward more generic terms would allow them to obtain a broad overview of the domain taxonomy. This can help them better understand the relationships between different classes and how they are organized in the hierarchy.

Exploring knowledge anchors (KA).

- Increasing the use of KA. These results suggest that increasing the use of KA has a positive impact on content diversity (CD). Therefore, an exploration strategy could prioritize the exploration of KA or entities that are directly linked to KA.
- Go to areas in the KG with higher KA rates. The results showed that the positive effect of KA on CD increased as users were exposed to higher rates of KA. Therefore, users can focus their exploration on areas with a higher concentration of KA, such as the top-level and middle levels in the class hierarchy.
- Use KA as a guide to explore related entities that are not directly linked to KA. This strategy can help users discover new and diverse content while remaining within the scope of their interests.

Overall characteristics of exploration strategies to support exploration experience are summarized in Table 9.

TABLE 9. Summary of characteristics of exploration strategies to support the categorize cognitive process.

Path parameter	Beginning of exploration path	Middle of exploration path	End of exploration path
top-level	√	-	√
middle-level	√	-	√
bottom-level	√	√	-
go-up	-	√	-
go-down	√	-	-
stay	-	-	-
high density	√	√	-
low density	-	-	-
KA	√	√	√
dead-end	-	-	-

C. HYBREDIZATION OF EXPLORATION STRATEGIES TO SUPPORT KNOWLEDGE UTILITY AND EXPLORATION EXPERIENCE

The exploration strategies used to support knowledge expansion and exploration experience were analyzed to identify common exploration strategies as follows.

Levels of generality in class hierarchy:

- Explore top-level and middle-level entities.
- Avoid exploring bottom-level entities through the exploration process, especially laypersons.
- Consider the density values of the top-level entities to identify important entities and their relationships.

Direction of exploration.

- Start the exploration at the top-level entities and then navigate down to the middle-level entities for a more in-depth exploration. Then, send users back to the top-level entities.

Exploring knowledge anchors (KA).

- Increasing the use of KA.
- Go to areas in the KG with higher KA rates, such as the top-level and middle-level of the class hierarchy.
- Use KA as a guide. Because KA are familiar concepts, users can use KA as a guide to explore related entities that they are unfamiliar with.

Density of entities.

- Initiate exploration with high-density entities at the top-level of the class hierarchy.
- Explore entities with high or low density in the middle of exploration.
- Explore dense entities at the end of exploration.

Dead ends:

- Avoid dead-end entities along the exploration path.

D. GENERALITY AND LIMITATIONS

The suggested exploration strategies are generic and can be applied individually or collectively across various domains represented as KGs. This wide applicability opens a diverse range of potential uses and applications. For example, exploration strategies can help users navigate through unfamiliar domains (e.g., drug industry) or partially familiar domains (e.g., cultural heritage or news), enabling them to gain insights and a deeper understanding of domain content. Furthermore, exploration strategies can facilitate exploration over large KGs where users may encounter overwhelming choices (e.g., health information, travel information and career options). By adopting exploration strategies, users can effectively explore a vast array of options, enabling them to locate relevant and valuable information more efficiently. In addition, the suggested strategies can be beneficial for users seeking to explore educational content and discover relevant resources and enhance their exploration experience. This can be extended further with a diversification strategy to suggest interesting concepts that are more likely to attract people to engage in exploration, and hence learn more about the domain.

However, there are limitations associated with the application of the suggested exploration strategies. For examples, the exploration strategies associated with levels of generality in the class hierarchy can only be applied to KGs with complex class hierarchies (i.e., class hierarchies with three levels of generality: bottom, middle and top levels), and cannot be applied to shallow class hierarchies (e.g., class hierarchies of depth = 1 or 2). Furthermore, exploration strategies that are based on the direction of exploration can be applied over directed graphs with the subsumption relationship rdfs:subClassOf as the main hierarchical relationship that links entities in the class hierarchy. Furthermore, the suggested exploration strategies for enhancing the exploration experience require pre-defined knowledge anchors (i.e., familiar concepts in the domain), which can be difficult to determine such generic concepts in specialized domains where lay users are not familiar with the domain concept.

VI. CONCLUSION

Knowledge graphs have become a valuable resource for exploration applications in various domains. However, the open-ended nature of exploratory search, combined with the abundance of exploration options provided by semantic links, can confuse and overwhelm users' exploration experience. Therefore, there is a need for appropriate exploration strategies to facilitate users' exploration of KGs. Previous attempts to facilitate exploration have focused on providing visual or textual interfaces; however, research in this area has since expanded to include research from diverse fields. This study aims to identify exploration strategies over KGs to support knowledge utility (i.e., increase users' domain knowledge) and exploration experience (i.e., provide users with a positive and pleasant feeling). To accomplish this goal, an experimental user study was conducted, involving lay users in the musical instrument domain, where they were presented with an exploration task and then allowed to freely explore musical instruments. Parameters related to exploration paths were identified and used to analyze the exploration patterns that users follow during their exploration. The key contribution of this study is the identification of path parameters that affect user exploration and the use of these parameters to analyze explorations paths to suggest exploration strategies over KGs to support knowledge utility and exploration experience. The outcome of this study contributes to the growing body of research on KG exploration and provides valuable insights into effective exploration strategies to support learning through search. The results of this study have practical implications for the design of user interfaces and systems that facilitate the exploration of KGs, particularly for users with limited domain expertise.

ACKNOWLEDGMENT

The authors are grateful to the participants in the experimental studies.

REFERENCES

[1] M. Lissandrini, D. Mottin, K. Hose, and T. B. Pedersen, "Knowledge graph exploration systems: Are we lost?" in *Proc. 12th Annu. Conf. Innov. Data Syst. Res. (CIDR)*. Honolulu, HI, USA: Chaminade Univ., Jan. 2022.

[2] L. Chang, M. Zhu, T. Gu, C. Bin, J. Qian, and J. Zhang, "Knowledge graph embedding by dynamic translation," *IEEE Access*, vol. 5, pp. 20898–20907, 2017.

[3] L. Zheng, S. Liu, Z. Song, and F. Dou, "Diversity-aware entity exploration on knowledge graph," *IEEE Access*, vol. 9, pp. 118782–118793, 2021.

[4] B. Abu-Salih, "Domain-specific knowledge graphs: A survey," *J. Netw. Comput. Appl.*, vol. 185, Jul. 2021, Art. no. 103076, doi: 10.1016/j.jnca.2021.103076.

[5] M. Lissandrini, T. B. Pedersen, K. Hose, and D. Mottin, "Knowledge graph exploration: Where are we and where are we going?" *ACM SIGWEB Newslett.*, vol. 2020, pp. 1–8, Jul. 2020.

[6] A. Soufan, I. Ruthven, and L. Azzopardi, "Searching the literature: An analysis of an exploratory search task," in *Proc. ACM SIGIR Conf. Hum. Inf. Interact. Retr.*, Mar. 2022, pp. 146–157.

[7] K. Huang, Q. Ye, J. Zhao, X. Zhao, H. Hu, and X. Zhou, "VINCENT: Towards efficient exploratory subgraph search in graph databases," *Proc. VLDB Endowment*, vol. 15, no. 12, pp. 3634–3637, Aug. 2022.

[8] T. Li, W. Wang, X. Li, T. Wang, X. Zhou, and M. Huang, "Embedding uncertain temporal knowledge graphs," *Mathematics*, vol. 11, no. 3, p. 775, Feb. 2023.

[9] M. Tibau, S. W. M. Siqueira, and B. P. Nunes, "Accounting for the knowledge gained during a web search: An empirical study on learning transfer indicators," *Library Inf. Sci. Res.*, vol. 45, no. 1, Jan. 2023, Art. no. 101222.

[10] T. Berners-Lee, Y. Chen, L. Chilton, D. Connolly, R. Dhanaraj, J. Hollenbach, A. Lerer, and D. Sheets, "Tabulator: Exploring and analyzing linked data on the semantic web," in *Proc. 3rd Int. Semantic Web User Interact. Workshop*, 2006, p. 159.

[11] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann, "DBpedia—A crystallization point for the web of data," *J. Web Semantics*, vol. 7, no. 3, pp. 154–165, Sep. 2009.

[12] G. Vassiliou, F. Alevizakis, N. Papadakis, and H. Kondylakis, "iSummary: Workload-based, personalized summaries for knowledge graphs," in *Proc. Eur. Semantic Web Conf.*, 2023, pp. 192–208.

[13] W. Wei, S. Zhao, and D. Zou, "Recommendation system: A survey and new perspectives," *World Sci. Annu. Rev. Artif. Intell.*, vol. 1, Jan. 2023, Art. no. 2330001.

[14] J. Cai, M. Li, Z. Jiang, E. Cho, Z. Chen, Y. Liu, X. Fan, and C. Guo, "KG-ECO: Knowledge graph enhanced entity correction for query rewriting," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.

[15] Y. Tu, X. Wang, R. Qiu, H.-W. Shen, M. Miller, J. Rao, S. Gao, P. R. Huber, A. D. Hollander, M. Lange, C. R. Garcia, and J. Stubbs, "An interactive knowledge and learning environment in smart foodsheds," *IEEE Comput. Graph. Appl.*, vol. 43, no. 3, pp. 36–47, May 2023.

[16] J. Peng, D. Xu, R. Lee, S. Xu, Y. Zhou, and K. Wang, "Expediting knowledge acquisition by a web framework for knowledge graph exploration and visualization (KGEV): Case studies on COVID-19 and human phenotype ontology," *BMC Med. Informat. Decis. Making*, vol. 22, no. S2, pp. 1–14, Dec. 2022.

[17] Y. Wang, A. Khan, X. Xu, J. Jin, Q. Hong, and T. Fu, "Aggregate queries on knowledge graphs: Fast approximation with semantic-aware sampling," in *Proc. IEEE 38th Int. Conf. Data Eng. (ICDE)*, May 2022, pp. 2914–2927.

[18] P. Vakkari, "Searching as learning: A systematization based on literature," *J. Inf. Sci.*, vol. 42, no. 1, pp. 7–18, Feb. 2016, doi: 10.1177/0165551515615833.

[19] J. Gwizdka, P. Hansen, C. Hauff, J. He, and N. Kando, "Search as learning (SAL) workshop 2016," in *Proc. 39th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, Jul. 2016, pp. 1249–1250, doi: 10.1145/2911451.2917766.

[20] S. Dietze and E. Kaldoudi, "Socio-semantic integration of educational resources—The case of the mEducator project," *Univ. Comput. Sci.*, vol. 19, no. 11, pp. 1543–1569, 2013.

[21] D. Thakker, V. Dimitrova, L. Lau, F. Yang-Turner, and D. Despotakis, "Assisting user browsing over linked data: Requirements elicitation with a user study," in *Proc. Int. Conf. Web Eng. (ICWE)*, in Lecture Notes in Computer Science, vol. 7977, 2013, pp. 376–383, doi: 10.1007/978-3-642-39200-9_31.

[22] M. Al-Tawil, D. Thakker, and V. Dimitrova, "Nudging to expand user's domain knowledge while exploring linked data," in *Proc. IESD@ISWC*, vol. 13, 2014, pp. 1–12.

[23] M. Al-Tawil, V. Dimitrova, and D. Thakker, "Using knowledge anchors to facilitate user exploration of data graphs," *Semantic Web*, vol. 11, no. 2, pp. 205–234, Feb. 2020, doi: 10.3233/SW-190347.

[24] D. P. Ausubel, "A subsumption theory of meaningful verbal learning and retention," *J. Gen. Psychol.*, vol. 66, no. 2, pp. 213–224, Apr. 1962, doi: 10.1080/00221309.1962.9711837.

[25] D. P. Ausubel, "The use of advance organizers in the learning and retention of meaningful verbal material," *J. Educ. Psychol.*, vol. 51, no. 5, pp. 267–272, Oct. 1960, doi: 10.1037/h0046669.

[26] D. G. Ausubel, "Cognitive structure and the facilitation of meaningful verbal learning," *J. Teacher Educ.*, vol. 14, no. 2, pp. 217–222, Jun. 1963, doi: 10.1177/002248716301400220.

[27] B. Abu-Salih, M. Al-Tawil, I. Aljarah, H. Faris, P. Wongthongtham, K. Y. Chan, and A. Beheshti, "Relational learning analysis of social politics using knowledge graph embedding," *Data Mining Knowl. Discovery*, vol. 35, no. 4, pp. 1497–1536, Jul. 2021.

[28] Y. Liang and M. C. Willemsen, "Promoting music exploration through personalized nudging in a genre exploration recommender," *Int. J. Hum.-Comput. Interact.*, vol. 39, no. 7, pp. 1495–1518, 2022.

[29] R. A. A. Principe, A. Maurino, M. Palmonari, M. Ciavotta, and B. Spahiu, "ABSTAT-HD: A scalable tool for profiling very large knowledge graphs," *VLDB J.*, vol. 31, pp. 851–876, Sep. 2022.

[30] K. Sakurai, R. Togo, T. Ogawa, and M. Haseyama, "Explainable artist recommendation based on reinforcement knowledge graph exploration," in *Proc. Int. Workshop Adv. Imag. Technol. (IWAIT)*, vol. 12177, May 2022, pp. 91–96.

[31] S. Kang, D. Lee, W. Kweon, and H. Yu, "Personalized knowledge distillation for recommender system," *Knowl.-Based Syst.*, vol. 239, Mar. 2022, Art. no. 107958.

[32] C.-W. H. Yuan, T.-W. Yu, J.-Y. Pan, and W.-C. Lin, "Interactive visual exploration of knowledge graphs with embedding-based guidance," in *Proc. Extended Abstr. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2023, pp. 1–8.

[33] S. Sellami and N. E. Zarour, "Keyword-based faceted search interface for knowledge graph construction and exploration," *Int. J. Web Inf. Syst.*, vol. 18, no. 5/6, pp. 453–486, Dec. 2022.

[34] Y. Diao, P. Guzewicz, I. Manolescu, and M. Mazuran, "Efficient exploration of interesting aggregates in RDF graphs," in *Proc. Int. Conf. Manage. Data*, Jun. 2021, pp. 392–404.

[35] M. Lissandrini, K. Hose, and T. B. Pedersen, "Example-driven exploratory analytics over knowledge graphs," in *Proc. 26th Int. Conf. Extending Database Technol. (EDBT)*, 2023, pp. 105–117.

[36] P. Organisciak, B. M. Schmidt, and J. S. Downie, "Giving shape to large digital libraries through exploratory data analysis," *J. Assoc. Inf. Sci. Technol.*, vol. 73, no. 2, pp. 317–332, Feb. 2022.

[37] Y. Yang, M. Li, J. Wang, W. Huang, and Y. Wang, "Entity recommendation with negative feedback memory networks for topic-oriented knowledge graph exploration," *IEEE Trans. Rel.*, vol. 71, no. 2, pp. 788–802, Jun. 2022.

[38] R. W. White and R. A. Roth, *Exploratory Search: Beyond the Query-Response Paradigm*. San Rafael, CA, USA: Morgan & Claypool, 2009, doi: 10.2200/S00174ED1V01Y200901ICR003.

[39] N. Belkin, "Anomalous states of knowledge as the basis of information retrieval," *Can. J. Inf. Sci.*, vol. 5, no. 1, pp. 133–143, 1980.

[40] G. Marchionini, "Exploratory search: From finding to understanding," *Commun. ACM*, vol. 49, no. 4, pp. 41–46, Apr. 2006, doi: 10.1145/1121949.1121979.

[41] D. Donato, L. Laura, S. Leonardi, and S. Millozzi, "The web as a graph," in *Proc. ACM SIGMOD-SIGACT-SIGART Symp. Principles Database Syst.*, 2000, pp. 1–10, doi: 10.1145/1189740.1189744.

[42] O. Lehmberg, R. Meusel, and C. Bizer, "Graph structure in the web," *Comput. Netw.*, vol. 33, pp. 309–320, Jun. 2000, doi: 10.1016/S1389-1286(00)00083-9.

[43] C. Bizer, T. Heath, and T. Berners-Lee, "Linked data—The story so far," *Int. J. Semantic Web Inf. Syst.*, vol. 5, no. 3, pp. 1–22, Jul. 2009, doi: 10.4018/jswis.2009081901.

[44] S. Xia, N. Anzum, S. Salihoglu, and J. Zhao, "KTabulator: Interactive ad hoc table creation using knowledge graphs," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2021, pp. 1–14.

[45] T. Aso, T. Amagasa, and H. Kitagawa, "A system for relation-oriented faceted search over knowledge bases," *Int. J. Web Inf. Syst.*, vol. 17, no. 6, pp. 698–713, Dec. 2021.

[46] N. Barret, A. Gauquier, J.-J. Law, and I. Manolescu, "PathWays: Entity-focused exploration of heterogeneous data graphs," in *Proc. 20th Eur. Semantic Web Conf.*, Hersonissos, Greece, May 2023.

[47] A. Indrawati, Z. Akbar, D. S. Rini, A. Yaman, Y. A. Kartika, and D. R. Saleh, "A knowledge graph exploration method with no prior knowledge," in *Proc. Int. Conf. Comput., Control, Informat. Appl.*, Nov. 2022, pp. 184–188, doi: 10.1145/3575882.3575918.

[48] N. Marie and F. Gandon, "Survey of linked data based exploration systems," in *Proc. IESD@ISWC*, no. 1, 2014, pp. 1–13. [Online]. Available: https://hal.inria.fr/hal-01057035/en

[49] J. M. Brunetti, R. García, and S. Auer, "From overview to facets and pivoting for interactive exploration of semantic web data," *Int. J. Semantic Web Inf. Syst.*, vol. 9, no. 1, pp. 1–20, Jan. 2013.

[50] A. G. Nuzzolese, V. Presutti, A. Gangemi, S. Peroni, and P. Ciancarini, "Aemoo: Linked data exploration based on knowledge patterns," *Semantic Web*, vol. 8, no. 1, pp. 87–112, Nov. 2016, doi: 10.3233/SW-160222.

[51] G. Troullinou, H. Kondylakis, E. Daskalaki, and D. Plexousakis, "Ontology understanding without tears: The summarization approach," *Semantic Web*, vol. 8, no. 6, pp. 797–815, Aug. 2017, doi: 10.3233/SW-170264.

[52] S. Ferré and A. Hermann, "Semantic search: Reconciling expressive querying and exploratory search," in *Proc. 10th Int. Semantic Web Conf.*, vol. 7031, 2011, pp. 177–192, doi: 10.1007/978-3-642-25073-6_12.

[53] G. Cheng, Y. Zhang, and Y. Qu, "Explass: Exploring associations between entities via top-K ontological patterns and facets," in *Proc. ISWC*, vol. 8797, 2014, pp. 422–437, doi: 10.1007/978-3-319-11915-1_27.

[54] B. Kules, R. Capra, M. Banta, and T. Sierra, "What do exploratory searchers look at in a faceted search interface?" in *Proc. 9th ACM/IEEE-CS Joint Conf. Digit. Libraries*, Jun. 2009, pp. 313–322, doi: 10.1145/1555400.1555452.

[55] T. Nunes and D. Schwabe, "Frameworks of information exploration—Towards the evaluation of exploration systems conceptual view of an exploration framework," in *Proc. IESD Workshop @ ISWC Conf.*, 2016.

[56] M. Al-Tawil, V. Dimitrova, D. Thakker, and B. Bennett, "Identifying knowledge anchors in a data graph," in *Proc. 27th ACM Conf. Hypertext Social Media*, Jul. 2016, pp. 189–194, doi: 10.1145/2914586.2914637.

[57] J. W. Tanaka and M. Taylor, "Object categories and expertise: Is the basic level in the eye of the beholder?" *Cogn. Psychol.*, vol. 23, pp. 457–482, Jul. 1991, doi: 10.1016/0010-0285(91)90016-H.

[58] G. Vega-Gorgojo, "LOD4Culture: Easy exploration of cultural heritage linked open data," *Semantic Web*, pp. 1–30, Jun. 2023.

[59] M. Sah and V. Wade, "Personalized concept-based search on the linked open data," *J. Web Semantics*, vol. 36, pp. 32–57, Jan. 2016, doi: 10.1016/j.websem.2015.11.004.

[60] G. A. Miller, "The magical number seven, plus or minus two: Some limits on our capacity for processing information," *Psychol. Rev.*, vol. 63, no. 2, p. 81, 1956.

[61] S. G. Hart and L. E. Staveland, "Development of NASA-TLX (task load index): Results of empirical and theoretical research," *Adv. Psychol.*, vol. 52, pp. 139–183, Apr. 1988, doi: 10.1016/S0166-4115(08)62386-9.

[62] S. C. Carr and B. Thompson, "The effects of prior knowledge and schema activation strategies on the inferential reading comprehension of children with and without learning disabilities," *Learn. Disability Quart.*, vol. 19, no. 1, pp. 48–61, Feb. 1996, doi: 10.2307/1511053.

[63] D. R. Krathwohl, "A revision of Bloom's taxonomy: An overview," *Theory Into Pract.*, vol. 41, no. 4, pp. 212–218, Nov. 2002.

[64] N. F. Noy and D. L. McGuinness. (2004). *Ontology Development 101: A Guide to Creating Your First Ontology*. [Online]. Available: http//protege.stanfordedu/publications

[65] A. Landherr, B. Friedl, and J. Heidemann, "A critical review of centrality measures in social networks," *Bus. Inf. Syst. Eng.*, vol. 2, no. 6, pp. 371–385, Dec. 2010, doi: 10.1007/s12599-010-0127-3.

MARWAN AL-TAWIL received the Ph.D. degree in computer science from the University of Leeds, Leeds, U.K., in 2018. He is an Assistant Professor with the Computer Information Systems Department, King Abdullah II School for Information Technology, The University of Jordan. He has practical experience in industry and has taught several courses on databases, knowledge management, and the semantic web. His current research interests include knowledge graph analysis and data exploration. His main focus is aiding users' exploration of knowledge graphs, particularly expanding their domain knowledge and enriching their exploration experience.

VANIA DIMITROVA is a Professor of human-centered artificial intelligence with the University of Leeds. She was the Co-Director of the Leeds Research Centre in Digital Learning and the Director of technology enhanced learning strategy with the Leeds Institute of Medical Education. She leads research activity on human-centered artificial intelligence, which builds intelligent systems, that help people make sense of data, take decisions in complex settings, expand their knowledge, learn from experience, and develop self-regulation skills. Her research explores the use of data and knowledge models to get insights into user-generated content, understand users and influence behavior, and capture knowledge and support information exploration. Her research is conducted in cross-disciplinary collaboration with researchers from medicine and health, engineering, social science, education and psychology, and actively involving end users. She is the President of the International AI in Education Society and the Co-Director of the UKRI Centre for Doctoral Training in AI for Medical Diagnosis and Care. She is an Associate Editor of the *International Journal of AI in Education* and *Frontiers of AI: AI for Human Learning and Behavior Change*. She was an Associate Editor of IEEE Transactions on Learning Technologies and a member of the editorial boards of the *Journal of Personalization* (UMUAI). She chaired the premier international conference on user modeling (ACM UMAP), key conferences in intelligent learning environments (AIED, ECTEL, and ICCE), and a series of international workshops on key topics related to intelligent mentoring, user modeling, social systems, and intelligent exploration.

DHAVALKUMAR THAKKER is a Professor of AI and the IoT with the University of Hull. He has over 15 years of working experience in the European Union (EU) and industrial projects, researched and delivered innovative solutions. He actively published in leading high impact factor journals, such as the *Semantic Web*, *Engineering Applications of Artificial Intelligence* (Elsevier), and *Transactions on Emerging Telecommunications Technologies*. His broad area of research interest and expertise is interdisciplinary focusing on the use of artificial intelligence (AI) and the Internet of Things (IoT) technologies for the betterment of society. His current and evolving research interests include exploring the role of AI and IoT technologies in the context of smart cities, digital health, and circular economy. His research has received multiple best paper awards (2019 at the tenth IEEE Conference on IoT, Big Data and AI for a Smart and Safe Future; and 2015 at 12th the European Semantic Web Conference). He regularly reviews for the Engineering and Physical Sciences Research Council (EPSRC) and the Natural Environment Research Council (NERC).

BILAL ABU-SALIH received the Ph.D. degree in information systems (with a focus on social big data analytics) from Curtin University. He is an Associate Professor with The University of Jordan and an Adjunct Professor with Curtin University. He worked on various cross-disciplinary funded research projects, that involve academic research, software development, and industrial implementation. His research interests include social big data, social trust, machine learning/data mining, knowledge graphs, NLP, and information retrieval.