The Future of Engineering Education
2024 Annual Conference & Exposition
Oregon Convention Center
Portland, OR . June 23 - 26, 2024
Paper ID #43699

WIP: A Knowledge Graph to Share and Discover High-Impact Practices and Support Decision-Making

Dr. Natalia Villanueva Rosales, The University of Texas at El Paso

Natalia Villanueva Rosales is an Associate Professor of Computer Science at The University of Texas at El Paso. Her work aims to improve the efficiency and effectiveness of the discovery, integration, and trust of data and models. Her approaches link human and machine knowledge to address societal-relevant problems in areas that require interdisciplinary research and collaborations across borders, such as sustainability of water resources and Smart Cities.

Dr. Ann C. Gates, The University of Texas at El Paso

Dr. Ann Quiroz Gates is the Senior Advisor to the Provost for STEM Initiatives and an AT&T Distinguished Professor in the Computer Science Department at The University of Texas at El Paso. Gates is the Executive Director of the NSF-funded Computing Alliance of Hispanic-Serving Institutions.

Lani Nicole Godenzi, The University of Texas at El Paso

Lani Godenzi is a Masters student at The University of Texas at El Paso pursuing a MS in Data and Information Science. She received her BS Chemical and Biological Engineering from Tufts University E'16. Her scholarly interests center bridging communication challenges between multidisciplinary teams to enable execution of technical projects that empower the next generation of engineering students.

Francisco Osuna, The University of Texas at El Paso

Francisco Osuna is a Software Engineer with a Master's in Computer Science from The University of Texas at El Paso with over 15 years of experience designing and developing robust software systems. Francisco spearheaded the design and development of Engage, a groundbreaking student success information management system that has been adopted by over 25 national academic institutions. This project alone has transformed academic administration by streamlining processes, reducing reporting overhead by an impressive 95%, and delivering on-time results through meticulous project management.

Angel U. Ortega, The University of Texas at El Paso

Angel U. Ortega is a Ph.D. candidate in the Computer Science Department at The University of Texas at El Paso. His area of research is in context-aware scientific natural language generation. Currently a Ph.D. Research Assistant, Angel also serves as a graduate advocate for CAHSI (The Computing Alliance of Hispanic-Serving Institutions), a mentor for the Coding Interview Club, and has interned at companies like Google, Redfin, and Bloomberg. Passionate about inclusive excellence, Angel shares CAHSI's mission to accelerate the progress of Hispanics in computing through his outreach, mentorship, and program development efforts.

Veronica A. Carrillo, The University of Texas at El Paso

Veronica A. Carrillo is a staff member at The University of Texas at El Paso who provides administrative and project management support. She earned a BBA in Management Information Systems from the University of Notre Dame. Beyond her professional responsibilities, Veronica is deeply passionate about civic engagement. Her involvement in her neighborhood association provides helpful insights into community building and engagement strategies that support the university's mission to enhance student success.

WIP: A Knowledge Graph to Share and Discover High-Impact Practices and Support Decision-Making

Abstract

This work-in-progress paper describes a Student Success Knowledge Graph (SSKG) that provides the foundation for documenting a shared understanding of student success and effective practices and discovering sources and experts of such efforts. The initiative contributes to a larger effort to create a culture of equity-minded, knowledge-driven decision-making. Using a participatory process that involves faculty, researchers, and administrators across campus, the research team documents information and knowledge centered on high-impact practices, student resources, student opportunities, identity and belonging in STEM, and components of inclusive excellence. Such a process involves diverse stakeholders in co-creating and validating the content of the SSKG and identifying relevant data sources. Through navigation of the SSKG using the custom-built interface, faculty, and administrators can discover practices used by departments and experts, adopters and experts associated with those practices, and supporting literature that informs the practices. This work aims to assist in knowledge-driven decision-making as chairs, faculty, and administrators seek to improve student retention and advancement in academic programs. This paper describes the creation of the SSKG and the implementation process, including the graphical interface and the question-answering that supports knowledge discovery.

1. Introduction

Systemic change for the success of a wide range of students requires orchestrated efforts across different entities within an institution and across institutions. To increase skills in data analysis for staff and faculty, our institution, The University of Texas at El Paso (UTEP), started an initiative to institutionalize the systematic use of data and knowledge to develop and implement initiatives designed to increase the success of students in Science, Technology, Engineering, and Mathematics (STEM) disciplines, particularly those from underserved communities. Theories of change note the complex set of factors that influence such outcomes [1] [2]. Our institution identified key progress metrics related to STEM programs and began diagnosing emergent issues that arose from data analysis. In addition, UTEP administers a student climate survey with a focus on non-cognitive and affective factors (e.g., belonging, identity, motivation), which are important in undergraduate student success [3]. To support decision-making, it is important to examine evidence-based practices, such as High Impact Practices (HIPs) [4], understand the specific context of those practices, and then decide on the appropriate approach to undertake. We have identified a wide variety of factors that can contribute to student success. Information and data related to these factors are often stored and maintained across various institutional systems, including individual and departmental data repositories. Integrating this information is a critical task for supporting informed decisions that can improve student retention and advancement in academic programs. Our team created a Student Success Knowledge Graph (SSKG) to address this challenge. Knowledge graphs enable the integration of data with ontologies as a backbone

[5]. Ontologies provide a formal representation of concepts and their relationships that enable the use of logic algorithms to derive new knowledge [6]. Knowledge graphs are commonly used to integrate heterogeneous data to enable easy navigation of these data for knowledge discovery and answering questions that require the use of contextual information. The National Science Foundation recently announced this year the creation of the Prototype Open Knowledge Network (Proto-OKN) that will host interconnected knowledge graphs and educational materials to support data-driven solutions [7].

Knowledge graphs have been used in diverse fields including medicine, cybersecurity, finance, news, and education among others [8]. To the best of our knowledge, related work in knowledge graphs focusing on education mostly centers on instructional subjects [9], [10], [11], [12], or educational material and resources [13]. Most of these approaches focus on defining the learning landscape, from topic to subject to course. Our approach aims to understand and describe the diverse practices and initiatives, curriculum among them, that might be factors of student success.

2. Enabling Knowledge Discovery for Student Success Data

This work presents the process of creating the SSKG and the capabilities of its first version (ver.) 1.0. The SSKG aims to provide a wide range of stakeholders with a means to navigate data related to student success, identify how data are related to each other, and discover new information and connections. The scenario provided in Fig. 1 elucidates how information from a knowledge graph could be used in the context of student success.

Scenario: Information Discovery in the Student Success Knowledge Graph (SSKG)

A chair from a STEM department has anecdotally observed the impact of undergraduate research experiences on developing students' skills in her department. She wants to encourage new faculty to involve undergraduate students in research and wants them to connect to other faculty that lead undergraduate research programs. She queries the knowledge graph and discovers faculty who are leading undergraduate research programs, and resources related to this practice. Through these resources, she learns that undergraduate research programs are considered a High Impact Practice that impacts STEM identity. She also finds a publication that provides good practices to build an undergraduate research program. The chair provides this information to new faculty and connects them to practitioners of this High Impact Practice to learn more about their institutional experience for this practice.

Figure 1. A scenario that illustrates the use of the SSKG for knowledge discovery.

3. A Concept Map to Represent Student Success Elements

To create the SSKG, we followed a bottom-up approach similar to the one proposed in [6] and commonly used to create ontologies.

3.1. Initial Concept Map and Competency Questions Development

The SSKG development was guided by competency questions [14] to be answered by the knowledge graph and information already available in our institution. These questions and concepts were identified by a constantly expanding group of administrators and faculty at UTEP. Many artifacts, including a concept map [15], were used to describe datasets and information. The creation of the concept map was also guided by competency questions. The research team followed an iterative process where the concept map and the competency questions were refined as more domain experts were involved in the process. The competency questions (CQ) used for the concept map presented in this paper are:

CQ1. Who are the practitioners or researchers on campus for each of the HIPs?

CQ2. What practices support (or inform) a particular outcome?

CQ3. What practices are used by a specific department?

CQ4. What research informs a particular HIP? What resources support a practice?

CQ5. What research informs a particular outcome? What resources support a particular outcome?

CQ6. What research addresses particular barriers?

CQ1 and CQ2 are used for illustrative purposes in the following sections. Guided by these competency questions, the first version of the concept map was collaboratively created by a core team of faculty and administrators that were part of an NSF-funded research group from the Provost Office and the Colleges of Education and Engineering (Award # 2122607). Through meetings focused on understanding the information available at our institution, the core team identified the main concepts related to the competency questions and their relationships. An iterative refinement process was followed, which involved additional meetings to further refine competency questions and validate the concept map as described in the next subsection.

3.2. Concept Map Refinement

Invitations were extended to individuals based on their specific areas of expertise and the relevance of the concepts and data under consideration to further validate the concept map with respect to accuracy and relevance. For instance, faculty members from the College of Education contributed to refining sections of the concept map related to `Barrier & Challenge to Student Success` and `K-12 STEM Preparation & Involvement`. Meanwhile, administrators well-versed in student success programming, for example, administrators and faculty from the Center for Community Engagement and student affairs offices undertook a thorough review of the `High Impact Practice` branch of the concept map. Additionally, administrators from the Provost Office provided valuable insights and guidance in the development of the `Inclusive Excellence and Institutional Culture` concept. Indeed, this participatory effort involved a diverse group of contributors, each offering their unique insights and knowledge to create a comprehensive concept map and identify sources of data relevant to the concepts being defined. The acknowledgment section lists contributors to the SSKG creation. As illustrated in Fig. 2, the main concepts around Student Success include: `Student Opportunity`, `High Impact Practice`, and `Student Resources` among others.

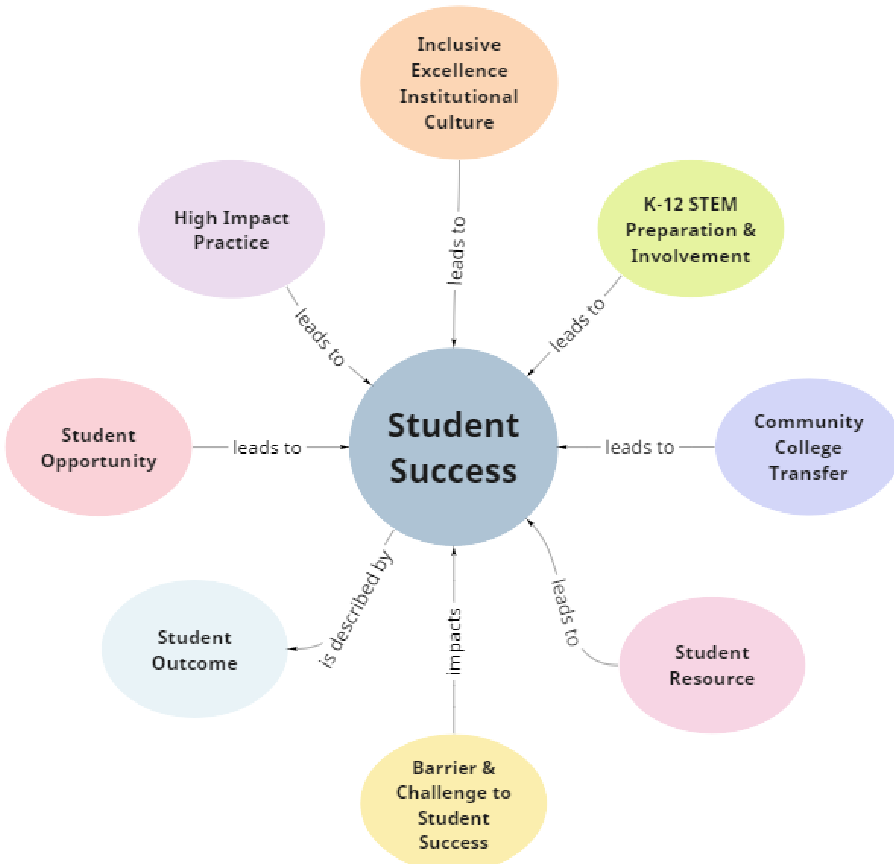

Figure 2. Main concepts related to student success in the SSKG ver. 1.0

In addition, concepts were linked to each other with a labeled, directed arrow to explicitly denote the relationship between them. For example, the concept `High Impact Practice` is linked to `Student Success` through the relationship `leadsTo` (Fig. 2) which describes that high impact practices are important factors in student success. Specific concepts are described in a hierarchy using the commonly used "is a" relationship. For example, we describe that `Student Employment` "is a" `High Impact Practice`, but not all `High Impact Practice` instances are of class `Student Employment` (Fig. 4). Other `High Impact Practice` include `Internship` and `First Year Experience`.

3.3. Glossary and Augmented Concept Map

To disambiguate the meaning of each concept and ensure mutual understanding of terms among stakeholders, a definition for all the concepts and an example of these concepts were included in a supplemental spreadsheet that was used as a **glossary**. The definitions were obtained from publications, organizations, and/or faculty members, and the source was captured in the glossary for provenance and attribution purposes. The glossary also captures additional details of data sources for each concept, e.g., if the data is subject to privacy policies or is public. Tables 1 and 2 are selected excerpts from the glossary to illustrate the type of information captured. Note that Table 1 includes the reference to the definition in this example for attribution purposes, however, the reference is not part of the glossary.

Table 1. Example of concept definitions that are included in the SSKG glossary.

Concept Name	Description/Definition	Source of the Definition	Data Provenance and Governance	Privacy policy	Example
High Impact Practice	"HIPs, based on evidence of significant educational benefits for students who participate in them—including and especially those from demographic groups historically underserved by higher education. These practices take many different forms, depending on learner characteristics and on institutional priorities and contexts." [16]	https://www.aacu.org/trending-topics/high-impact	Student Involvement Tool, Faculty & Staff	Public	Building Scholars, U-RISE Program, COURI, iLink REU.

Table 2. Example of a concept example (instance) from the Glossary

Concept	Example	Faculty or Expert	Example URL
Undergraduate Research Program	iLink REU	Natalia Villanueva Rosales	https://ilink.cybershare.utep.edu/

The research team created an augmented concept map to include additional classes such as publications, faculty, organizations, and applicable relations that are needed to answer the competency questions but are not core concepts in the SSKG scope. This augmented concept map was used only by the internal team as the complexity of the links and relationships would hinder the readability of the map for discussion among faculty and staff working on specific concepts on the initial concept map. For example, the explicit definition of `Faculty isAffiliatedWith Department` is included in the augmented concept map as it is needed for inference by the reasoner (see Section 6.1 and 6.2), but does not add to discussions with administrators about HIPs. The augmented concept map was implemented as an ontology described in the following section.

4. From a Concept Map to the Student Success Knowledge Graph

The SSKG was implemented as an ontology in the OWL format [17] using the augmented concept map and glossary described in the previous section. Using the third-party Protégé tool [18], each node from the concept map was created as an OWL Class. The relationships in the concept map, depicted as directed arrows, were represented as OWL Object Properties, e.g., `SSKG:impacts` (Fig. 3c). The special relationship "is a", describes the subclass hierarchy and was represented with the `rdfs:subClassOf` property from the RDF language [19]. The definition of each class from the previously created glossary and the source for this definition was described using annotation properties to keep track of the provenance and provide attributions for the concepts and data. The current version of the knowledge graph includes data provided by collaborators that is publicly available. For example, publications included information about the author, title, and DOI. Widely used vocabularies were used when possible, e.g., the `sch:ScholarlyArticle` from Schema.org [20]. Similarly, Dublin Core vocabulary [21] properties, e.g., `dc:source,` were used to describe the source of information for attribution and provenance purposes. We followed ontology design good practices [14], including reusing vocabulary when available to promote interoperability and exchange of data with other applications. An excerpt of the ontology description of the class `High Impact Practice` is presented in Fig. 3 (a) using Manchester Syntax [22]. The prefix SSKG denotes the terminology developed in the SSKG.

(a)
```
Class: SSKG:HighImpactPractice
    Annotations:
        dc:source <https://www.aacu.org/trending-topics/high-impact>,
        dc:description "HIPs, based on evidence of significant
educational benefits for students who participate in them—including
and especially those from demographic groups historically underserved
by higher education. These practices take many different forms,
depending on learner characteristics and on institutional priorities
and contexts."^^rdfs:Literal
    SubClassOf:
        SSKG:Impacts some SSKG:Outcome,
        SSKG:LeadsTo some SSKG:StudentSuccess
```

(b)
```
Individual: SSKG:iLink_REU
    Annotations:
        dc:source < https://ilink.cybershare.utep.edu/>
        dc:description "The iLink group provides research and training
opportunities for undergraduate and graduate students through multiple
research projects funded by different agencies, organizations and
industry partners."
    Types:
        SSKG:UGResearchProgram
```

(c)
```
ObjectProperty: SSKG:impacts
    SubPropertyChain:
        SSKG:ReportedIn o SSKG:DescribesOutcome
```

Figure 3. (a) Description of the SSKG `High Impact Practice` *class, (b) an example of an* `Undergraduate Research Program` *instance individual, and (c) an Object Property role chain in Manchester Syntax.*

Several versions of the knowledge graph were created through an iterative refinement process aligned with the refinement of the concept map, competency questions, and analysis of data sources. Hierarchies were created starting with concepts for which data was available in our organization. Illustrative examples for most HIPs were selected from the data sources discussed in the following section and added as OWL Individuals in the first version of the ontology, SSKG ver. 1.0.

The current SSKG ontology contains 128 classes, and 39 object properties.

5. Populating the Knowledge Graph

The team developed a methodology for data integration to populate the SSKG. The scope for the SSKG ver 1.0 includes information about three departments from two different colleges in our institution, engineering and science. Information to populate the `High Impact Practice` subgraph was mostly obtained by leveraging the UTEP Engage tool that contains a database of historical faculty and student involvement in HIPs at our institution. This tool produces spreadsheet reports, thus, populating the ontology involved a manual process using the Protégé tool. This knowledge base provided individual instances of HIPs such as faculty-led workshops, undergraduate research opportunities, internships, and courses where HIPs were used to enhance student engagement in the selected departments. For the first version of the SSKG, student information and other non-public information were excluded. We plan to include this information in a subsequent version once the policies for aggregation and anonymization (if applicable) are applied to new data.

Most of the instances of HIPs obtained from our student involvement tool were directly mapped to classes in our SSKG ontology. This information was provided directly by faculty and staff through the UTEP Engage tool. Fig. 3(b) shows the definition of the instance `SSKG:iLink_REU` in Manchester Syntax. Relationships were also identified from the tabular data provided, for example, faculty member(s) leading an initiative were linked to the initiatives through the property `SSKG:isResponsibleFor.` Fig. 3(c) shows the definition of the property `SSKG:impacts.` During the ontology design process, we observed that faculty members who are practitioners of HIPs were acting as connectors between HIPs, outcomes, and other concepts. Faculty members were added to the knowledge graph using the text of their biography published on their departmental website or faculty directory to generate the corresponding instances.

The SSKG currently contains 136 individual instances.

6. Answering Questions with the Student Success Knowledge Graph

To enable question answering from the knowledge graph, the competency questions were mapped to SPARQL [23] queries. SPARQL queries enable the retrieval of data that matches a subgraph

(Fig. 5) from the SSKG knowledge graph. Our first step towards question-answering in SSKG involved using the HERMIT reasoner [24] in Protégé and the Snap SPARQL [25] plugin.

Fig. 4 shows an excerpt of the concept map that includes two levels of concepts for the HIPs and `Student Outcome`. For example, specific HIPs included in the SSKG concept map are `Internship`, `Capstone Course`, and `Research and Scholarly Activity`. Each HIP is further refined to identify the variations of practices, which allows faculty and staff to adopt practices and identify existing workshops, activities, or other resources that can assist in adopting or adapting HIPs. Such data, coupled with other data like student outcome, contribute to creating a culture of equity-minded, data-driven decision-making.

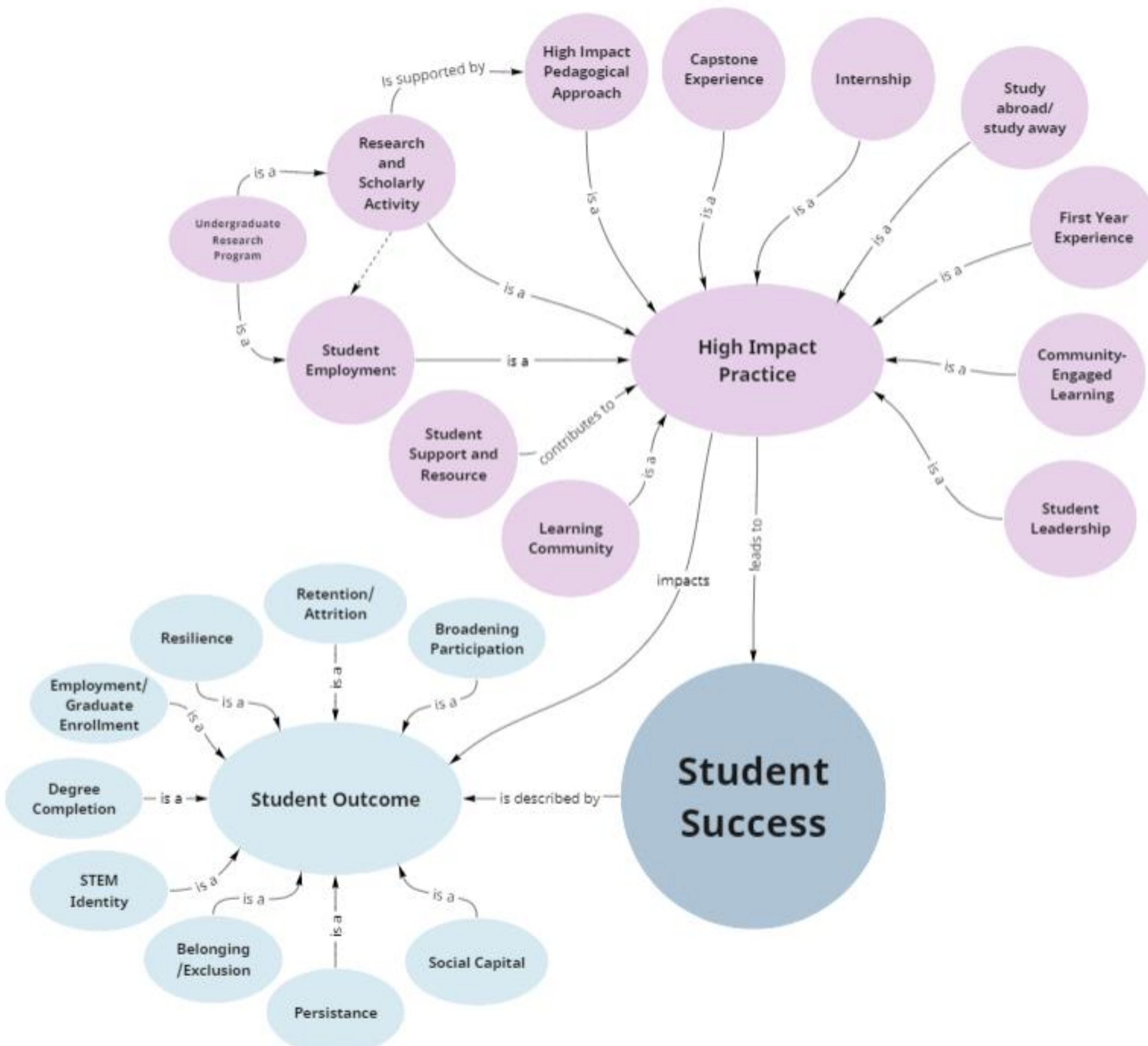

Figure 4. Excerpt of the SSKG concept map with level 2 expansion for the concepts `High Impact Practice` and `Student Outcome`.

6.1. Answering CQ1: Query 1

Related to our scenario in Section 1, a department chair in our institution wants to learn more about HIPs in other departments, the faculty that are responsible for those practices, and the department with which they are affiliated. This query is relevant to CQ1 presented in Section 3. Query 1 can be graphically represented by the subgraph in Fig. 5.

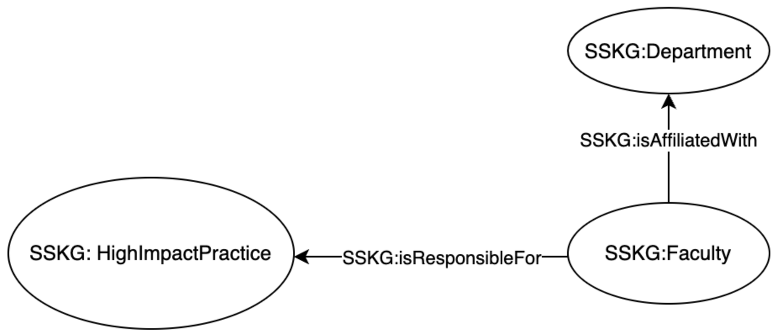

Figure 5. Graph representation of Query 1 that provides information relevant to CQ1, i.e., faculty responsible for HIPs and their affiliation.

To answer Query 1, the SPARQL query presented in Figure 6 was created. Note that we use the DISTINCT modifier to retrieve only different HIPs and eliminate duplicates.

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX SSKG: <http://ontology.cybershare.utep.edu/StudentSuccess/#>

SELECT DISTINCT ?HIP ?Practitioner ?Department WHERE {

     ?Practitioner SSKG:isResponsibleFor ?HIP.
     ?Practitioner SSKG:isAffiliatedWith ?Department.
     ?Practitioner rdf:type              SSKG:Faculty.
     ?HIP          rdf:type              SSKG:HighImpactPractice.
     ?Department   rdf:type              SSKG:Department.

}
```

Figure 6. Query 1 SPARQL representation that retrieves practitioners of a HIP and their corresponding department.

Query 1 results. The results provide information about faculty who are responsible for HIPs and their department. For example, `SSKG:Natalia_Villanueva_Rosales` is responsible for `SSKG:iLink_REU` which is an instance of an `SSKG:UGResearchProgram`. A `SSKG:UGResearchProgram` is a subclass of a `SSKG:ResearchAndScholaryActivity` which in turn is a subclass of a

SSKG:HighImpactPractice. Because of the rdfs:subClassOf property, a reasoner is able to infer that the instance of a class is also an instance of its superclass. Therefore, the SSKG:iLink_REU is classified as a SSKG:HighImpactPractice. In addition, Query 1 retrieves the Department_of_Computer_Science to which SSKG:Natalia_Villanueva_Rosales belongs through the property SSKG:isAffiliatedWith.

6.2. Answering CQ2: Query 2

Suppose that the departmental chair also wants to learn more about the outcomes related to HIPs. This is also related to our scenario in Section 1. Query 2 can be represented graphically in the subgraph of Fig. 7. Note that there is not an explicit relation between SSKG:HighImpactPractice and SSKG:Outcome. However, our SSKG ontology contains the role chain SSKG:reportedIn o SSKG:describesOutcome -> rdfs:subPropertyOf SSKG:impacts which indicates that, if the pattern of the property SSKG:reportedIn followed by the property SSKG:describesOutcome is found, the new property SSKG:impacts can be created at query time or when inferences are pre-computed. Note that this inference is only used when results are retrieved, and not stored (i.e. asserted), from the knowledge graph. In Fig. 7 we represent an inferred property with dotted lines graphically. Query 2 illustrates the ability to retrieve information that is not explicitly stated, which is a key difference between knowledge graphs and relational databases.

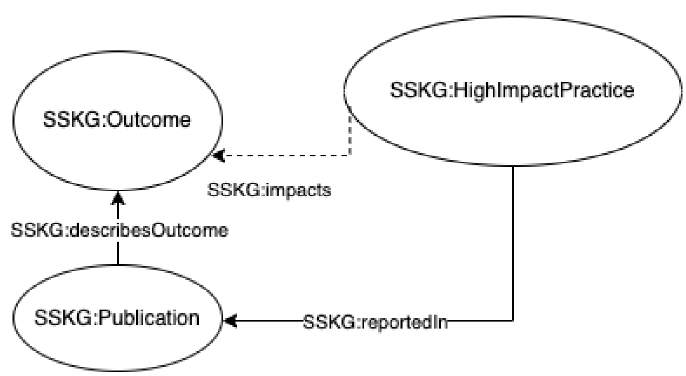

Figure 7. Graphical representation of concepts related to Query 2 that retrieves HIPs and the outcomes they impact.

The SPARQL representation of Query 2 is shown in Fig 8.

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX SSKG: <http://ontology.cybershare.utep.edu/StudentSuccess/#>

SELECT DISTINCT ?HIP ?Outcome WHERE {
      ?HIP     SSKG:impacts ?Outcome.
      ?HIP     rdf:type     SSKG:HighImpactPractice.
      ?Outcome rdf:type     SSKG:Outcome.
}
```

Figure 8. Query 2 SPARQL representation that retrieves HIPs and the student outcomes they impact.

Query 2 results. The results provide information about HIPs and the outcomes they represent. In the results, we find `SSKG:UndergraduateResearch,` which is described in `SSKG:Publication49` (i.e., a paper), to be of type `SSKG:Resource`. This resource provides evidence (i.e. reports) that `SSKG:UndergraduateResearch` has an impact on `SSKG:Belonging` (classified as a student `SSKG:Outcome`). Due to the role chain previously described, the `SSKG:impacts` relation is generated by the reasoner.

Query 1 and Query 2 illustrate how a user can retrieve information from the knowledge graph that is explicitly stated (i.e., asserted) or that can be inferred with the use of a reasoner which is one of the capabilities of knowledge graphs that are not available in regular databases.

7. Enabling Knowledge Discovery and Information Retrieval from the Student Success Knowledge Graph

A core challenge in visualizing information in the SSKG lies in managing the inherent complexity of displaying a large number of concepts and their relations while preventing information overload for users. To address this challenge, our approach includes the creation of a Web-based interface.

7.1. User Interface Development

The Web-based interface has two distinct features: Textual Filtering and Graph Exploration. This design addresses the challenge of knowledge graph visualization by combining the efficiency of textual filtering with the exploration features of graph visualization. It strikes a balance between rapid information retrieval and comprehensive exploration, reducing information overload and empowering users to navigate and comprehend complex ontological datasets.

7.1.1. Textual Filtering Feature

The first feature of our interface enables users to perform a textual filter, a search form that offers several advantages:

Efficient Information Retrieval: Users can search through keywords, a common approach for accessing information within large datasets.

Reduced Information Overload: By presenting users with a limited set of search results, we mitigate the risk of overwhelming them with large amounts of data. The interface further deconstructs the knowledge graph into three main search areas of interest: *High Impact Practice, Practitioner,* and *Resource*. Each interface enables the user to view results from a different perspective.

Fig. 9 displays the initial prototype of the HIPs textual filtering interface. Users can choose a department and practice type, which yields a list of all HIPs within the SSKG. This list includes related information such as HIPs' names, practitioners, types, associated departments, and descriptions. Note that the *High Impact Practice* tab of the Web-based interface displays the information retrieved in Query 1 and other relevant information. In this scenario, all the HIPs associated with the Computer Science department that are categorized under *High Impact Practice (*e.g., *Student Employment, Research and Scholarly Activity)* are displayed to the user. The user can filter by selecting the practice type(s) of interest, as well as navigate through the interlinked concepts yielded from the results.

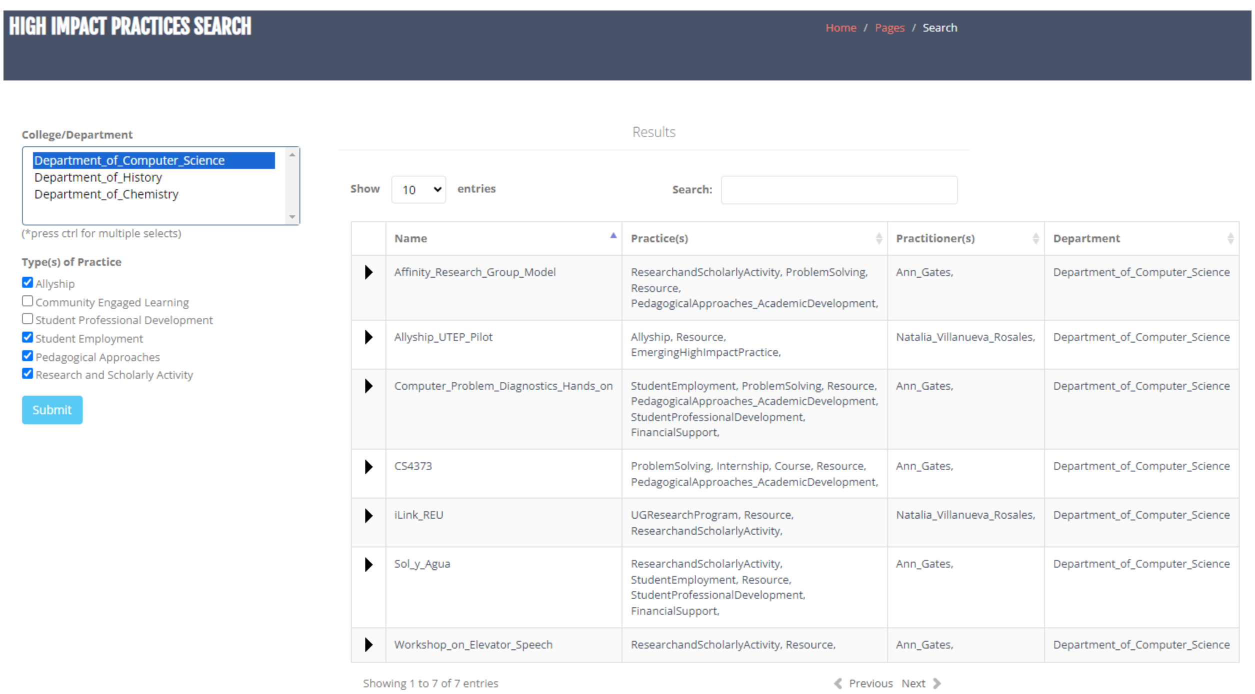

Figure 9. Initial prototype of the HIPs textual filtering interface.

7.1.2. Visualization Exploration – Work In Progress

The second feature of our interface leverages 3D Force Graph, a dynamic and interactive open-source framework for visualizing and exploring graph data [26]. This feature offers the following benefits:

Contextual Exploration: Users can interactively explore the SSKG graph structure through a visual representation. Users can also expand nodes of interest and examine relationships to discover more context on data connections.

Visual Representation: 3D Force Graph offers a three-dimensional representation of the knowledge graph, which enriches users' structural understanding of concepts, instances, and relationships, including hierarchies.

Contextual Filtering: Dynamic filtering options are introduced based on the user's current exploration of the graph.

Interactive Experience: The interface allows users to interact with the visualization by zooming, panning, and clicking on nodes for detailed information. This interactivity aims to foster user engagement and facilitate a user-friendly exploration experience also allowing for data discovery.

Complementary Approaches: Our hybrid interface integrates text-based searching and visual exploration, enabling users to employ both methods as needed. This flexibility accommodates diverse user preferences and query types.

Fig. 10 shows an example of the exploratory visualization derived from interconnected knowledge graph nodes that emerge when a user selects a HIP categorized under `Affinity_Research_Group_Model`. This visualization reveals the related resources, programs, individuals, types, outcomes, and organizations, providing a deeper understanding of the contextual relationships surrounding that concept. To minimize information overload and improve readability, relationship names are hidden. When a user hovers over a link, the corresponding name is displayed in red. This contextual highlighting helps the user focus on specific details without being overwhelmed by too much information at once. When right-clicking a node, the user is presented with a pop-up overlay box showing contextual information about that particular node. Users can also pan, rotate, and zoom in to focus on specific areas of the knowledge graph. The feature enables a more flexible exploration of resources and their relationships, e.g., Fig. 10 shows partial information retrieval based on Query 2. The visualization allows the research team to view and refine the SSKG.

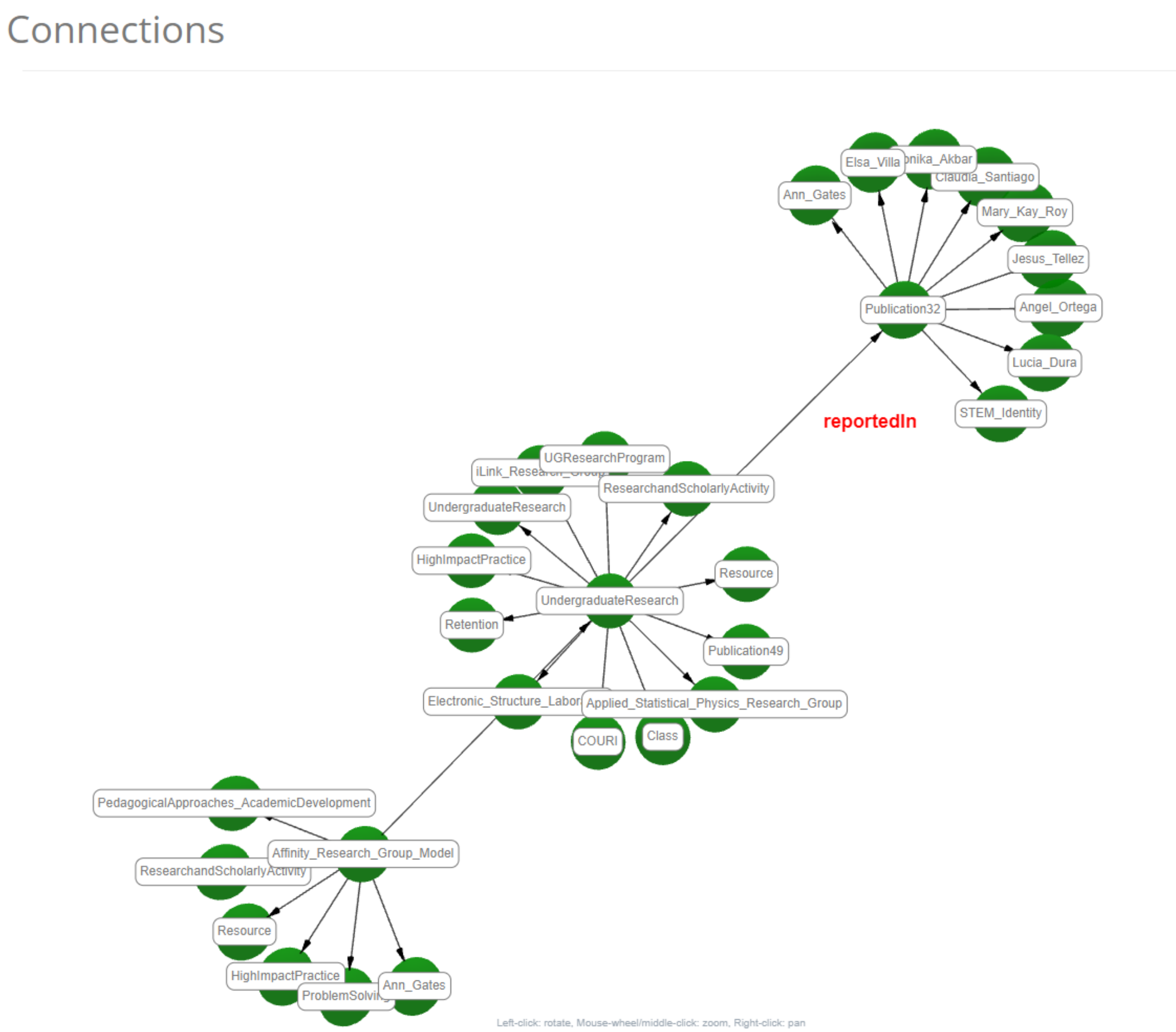

Figure 10. Example of an exploratory visualization of Knowledge Graph nodes in the interface prototype.

The visual exploration feature of the SSKG is currently a proof-of-concept undergoing iterative revisions to test and validate the usability needed before finalizing all features. Our ongoing work involves fully connecting all cases dynamically with the SSKG through SPARQL queries.

This interface prototype is currently available through our institutional servers and is only visible to members of our institution. We expect to make this interface publicly available in the future.

8. Conclusions and Future Work

In this paper, we present the SSKG that integrates information and domain expertise (i.e., concepts and relationships) about factors that are related to students' success in our institution. This participatory, collaborative effort involved different stakeholders who provided domain expertise and data to populate the ontology as the backbone of the SSKG. The integration of decoupled, heterogeneous data included an initial validation of the data sources with faculty and administrators to ensure that the data and its relationships were properly represented. The use of the Engage tool that captures student and faculty activities facilitated this process. However, this was still a resource-intensive process as additional data was manually integrated. We used established practices for the development of the ontology that is at the backbone of the SSKG, including the reuse of relevant vocabularies, when available. Our current work includes creating a workflow for the automation of the data integration process. We are working towards a self-sustaining and evolving knowledge graph.

Our interface prototype allows users to navigate data intuitively, and efforts will continue to refine the interface through usability tests involving a diverse group of stakeholders. The SSKG enables data discovery and supports informed decision-making related to student success through guided navigation of relevant information and knowledge from disparate sources. The SSKG ver. 1.0 enables question-answering related to student success factors that leverage logical, formal descriptions and knowledge inference. This is a key difference between the use of knowledge graphs and relational databases.

The SSKG can also provide information and knowledge for additional data analysis or data integration/augmentation that can integrate machine learning methods to discover patterns and support student success. Additionally, new AI technologies, including Large Language Models, can be leveraged to extend knowledge graphs [27], e.g., the SSKG, or be used to inform contextualized language models [28].

The SSKG and interface can support the informed decision-making of faculty and administrators interested in learning about and adopting HIPs that impact student success through evidence-based approaches. This framework and process can be adopted by other organizations that have technical resources and the ability to collect and integrate data.

9. Acknowledgements

The authors would like to thank the following contributors to the development and refinement of the concept map and competency questions, knowledge graph, and/or providing relevant data for the population of the SSKG: Toni Blum, Dania Brandford-Calvo, Christina Convertino, Erick Dominguez, Lourdes Echegoyen, Virginia Fraire, Issac Frausto Hernandez, Azuri Gonzalez, Jennifer Lujan, Tami Keating, Jessica Rivera, Louie Rodriguez, Deana Pennington, Elsa Villa, John Wiebe, Lucina Zarate. The authors would also like to thank the reviewers who contributed to improving the paper's quality.

This material is based upon work supported by the National Science Foundation under Grant No. 2122607 Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[1] A. S. Bryk, “2014 AERA Distinguished Lecture: Accelerating How We Learn to Improve,” *Educational Researcher*, vol. 44, no. 9, pp. 467–477, 2015, doi: 10.3102/0013189X15621543.

[2] A. Kezar, “Higher Education Change and Social Networks: A Review of Research,” *The Journal of Higher Education*, vol. 85, pp. 91–125, Jan. 2014, doi: 10.1353/jhe.2014.0003.

[3] M. Scheidt *et al.*, “Engineering students’ noncognitive and affective factors: Group differences from cluster analysis,” *Journal of Engineering Education*, vol. 110, no. 2, pp. 343–370, 2021, doi: https://doi.org/10.1002/jee.20386.

[4] G. D. Kuh, “Excerpt from high-impact educational practices: What they are, who has access to them, and why they matter,” *Association of American Colleges and Universities*, vol. 14, no. 3, pp. 28–29, 2008.

[5] L. Ehrlinger and W. Wöß, “Towards a Definition of Knowledge Graphs,” *SEMANTICS 2016: Posters and Demos Track*, Sep. 2016, [Online]. Available: https://ceur-ws.org/Vol-1695/paper4.pdf

[6] M. Uschold and M. Gruninger, “Ontologies: principles, methods and applications,” *The Knowledge Engineering Review*, vol. 11, no. 2, pp. 93–136, Jun. 1996, doi: 10.1017/S0269888900007797.

[7] U.S. National Science Foundation, “NSF invests $26.7 million in building the first-ever prototype open knowledge network.” [Online]. Available: https://new.nsf.gov/tip/updates/nsf-invests-first-ever-prototype-open-knowledge-network

[8] X. Zou, “A Survey on Application of Knowledge Graph,” *Journal of Physics: Conference Series*, vol. 1487, no. 1, p. 012016, Mar. 2020, doi: 10.1088/1742-6596/1487/1/012016.

[9] P. Chen, Y. Lu, V. W. Zheng, X. Chen, and X. Li, “An automatic knowledge graph construction system for K-12 education,” in *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*, in L@S ’18. New York, NY, USA: Association for Computing Machinery, 2018. doi: 10.1145/3231644.3231698.

[10] P. Chen, Y. Lu, V. W. Zheng, X. Chen, and B. Yang, “KnowEdu: A System to Construct Knowledge Graph for Education,” *IEEE Access*, vol. 6, pp. 31553–31563, 2018, doi: 10.1109/ACCESS.2018.2839607.

[11] M. Rizun, “Knowledge graph application in education: a literature review,” *Acta Universitatis Lodziensis. Folia Oeconomica*, vol. 3, no. 342, pp. 7–19, 2019.

[12] Y. Qin, H. Cao, and L. Xue, “Research and Application of Knowledge Graph in Teaching: Take the database course as an example,” *Journal of Physics: Conference Series*, vol. 1607, no. 1, p. 012127, Aug. 2020, doi: 10.1088/1742-6596/1607/1/012127.

[13] N. Li, Q. Shen, R. Song, Y. Chi, and H. Xu, “MEduKG: A Deep-Learning-Based Approach for Multi-Modal Educational Knowledge Graph Construction,” *Information*, vol. 13, no. 2, p. 91, Feb. 2022, doi: 10.3390/info13020091.

[14] N. F. Noy and D. L. McGuinness, “Ontology development 101: A guide to creating your first ontology.” Stanford knowledge systems laboratory technical report KSL-01-05 and …, 2001.

[15] R. R. Starr and J. M. P. de Oliveira, “Concept maps as the first step in an ontology construction method,” *Information Systems*, vol. 38, no. 5, pp. 771–783, 2013, doi: https://doi.org/10.1016/j.is.2012.05.010.

[16] American Association of Colleges and Universities, “High Impact Practices,” AAC&U. Accessed: May 01, 2024. [Online]. Available: https://www.aacu.org/trending-topics/high-impact

[17] W3C OWL Working Group, “OWL 2 Web Ontology Language Document Overview (Second Edition).” Accessed: Jul. 10, 2017. [Online]. Available: https://www.w3.org/TR/owl2-overview/

[18] M. A. Musen, “The protégé project: a look back and a look forward,” *AI Matters*, vol. 1, no. 4, pp. 4–12, 2015, doi: 10.1145/2757001.2757003.

[19] R. Cyganiak, D. Wood, and M. Lanthaler, “RDF 1.1 concepts and abstract syntax,” *World Wide Web Consortium, Working Draft WD-rdf11-concepts-20130723*, 2013.

[20] R. V. Guha, D. Brickley, and S. Macbeth, “Schema.Org: Evolution of Structured Data on the Web,” *Commun. ACM*, vol. 59, no. 2, pp. 44–51, Jan. 2016, doi: 10.1145/2844544.

[21] S. Weibel, J. Kunze, C. Lagoze, and M. Wolf, “Dublin core metadata for resource discovery,” 1998.

[22] M. Horridge and P. F. Patel-Schneider, “OWL 2 Web Ontology Language Manchester Syntax (Second Edition),” 2012, [Online]. Available: http://www.w3.org/TR/2012/NOTE-owl2-manchester-syntax-20121211/

[23] E. Prud’Hommeaux, and A. Seaborne, “SPARQL query language for RDF,” *W3C recommendation*, vol. 15, 2008.

[24] R. Shearer, B. Motik, and I. Horrocks, “HermiT: A Highly-Efficient OWL Reasoner.,” in *OWLED*, 2008, p. 91.

[25] M. Horridge and M. A. Musen, “Snap-SPARQL: A Java Framework for Working with SPARQL and OWL,” in *OWL: Experiences and Directions*, 2015. [Online]. Available: https://api.semanticscholar.org/CorpusID:37662788

[26] vasturiano, “3d-force-graph (Version 1.73),” *GitHub*. [Online]. Available: https://github.com/vasturiano/3d-force-graph

[27] J. Wang *et al.*, “TechGPT-2.0: A large language model project to solve the task of knowledge graph construction,” 2024, doi: 10.48550/ARXIV.2401.04507.

[28] H. Fei, Y. Ren, Y. Zhang, D. Ji, and X. Liang, “Enriching contextualized language model from knowledge graph for biomedical information extraction,” *Briefings in Bioinformatics*, vol. 22, no. 3, p. bbaa110, Jun. 2020, doi: 10.1093/bib/bbaa110.