Article

Ontology-Based Linked Data to Support Decision-Making within Universities

Ghadeer Ashour [1,2,*], **Ahmed Al-Dubai [1]**, **Imed Romdhani [1]** and **Daniyal Alghazzawi [3]**

1   School of Computing, Edinburgh Napier University, Edinburgh EH10 5DT, UK
2   Applied College, King Abdulaziz University, Jeddah 21589, Saudi Arabia
3   Department of Information Systems, King Abdulaziz University, Jeddah 21589, Saudi Arabia
*   Correspondence: gaashor@kau.edu.sa

Abstract: In recent years, educational institutions have worked hard to automate their work using more trending technologies that prove the success in supporting decision-making processes. Most of the decisions in educational institutions rely on rating the academic research profiles of their staff. An enormous amount of scholarly data is produced continuously by online libraries that contain data about publications, citations, and research activities. This kind of data can change the accuracy of the academic decisions, if linked with the local data of universities. The linked data technique in this study is applied to generate a link between university semantic data and a scientific knowledge graph, to enrich the local data and improve academic decisions. As a proof of concept, a case study was conducted to allocate the best academic staff to teach a course regarding their profile, including research records. Further, the resulting data are available to be reused in the future for different purposes in the academic domain. Finally, we compared the results of this link with previous work, as evidence of the accuracy of leveraging this technology to improve decisions within universities.

1. Introduction

In the last few years, higher education institutions (HEIs), such as universities, are increasingly using more modern technologies to automate different activities and improve the quality of their data. One of these technologies is representing academic data semantically in RDF format. Research, employment, and decision-making are examples of the challenging activities that higher education (HE) entails. Due to the nature of and frequent increase in academic data, semantic representation succeeds in solving several challenges in the educational domain. Although semantics have proven effective in many aspects, some shortages were diagnosed, such as dealing with missing information and the continuous update of data.

On the other hand, HEIs, such as universities, are increasingly using linked data (LD) to make public information (academic programs, research outputs, facilities, etc.) available as linked data on the Web. This trend opens the opportunities to use these data to automate the accomplishment of main processes within several institutions. Digital libraries are one of the institutions that use LD to publish scientific data and make them available to be used freely by others.

This research examines the outcome of a linked data creation cycle in the context of academic scientific research. It relies on Saudi university quality accreditation regulations. The study investigates the added value of leveraging the semantic technology of linked data in decision-making to produce accurate results for different tasks. The conducted scenario is applied to the local data of the Faculty of Computing and Information Technology

(FCIT) within King Abdulaziz University (KAU) to propose the engagement of linked data techniques. It aims to automatically enrich the local data and support the decision-making process within universities, such as course–teacher assignments.

The following partial objectives are established to attain the goal of this study:

1.　Identify a use case and reveal the main objectives of LD.
2.　Present a methodology to generate the link between university ontology and external academic staff scholarly data.
3.　Conduct a survey to investigate the elements that most affect the course–teacher assignment process.
4.　Demonstrate SPARQL queries for testing the resulting dataset to illustrate the success of using LD technology, by presenting SPARQL queries according to the most elements that affect the decision.
5.　Compare the results with previous work that uses semantic technology only to solve the same problem.

This work is organized as follows; Section 2 identifies this study's motivation and background. Then, Section 3 discusses the related works that used the LD technique in education. After that, the applied methodology to generate LD is illustrated in Section 4, followed by the results and discussion presented in Section 5. Finally, the conclusion is demonstrated in Section 6.

2. Motivation and Background

2.1. Challenges in Higher Education

Since it creates workers with a variety of specialties for all institutions, HE is considered the foundation for constructing the future globally. Therefore, supporting it with all the supplements that ensure effective performance is essential. Research, employment, and decision-making are all significant components of HE, in addition to teaching. Most HEIs, especially universities, improve their traditional processes of managing these components and solve the challenges related to them.

The challenge of allocating the best possible academic teacher to teach a new course is addressed in our previous studies [1,2]. It is one of the most common challenges universities are facing continuously under their decision-making processes. It depends on matching course contents with academic resource qualifications. A study [2] proposed an educational ontology to replace the traditional processes that heads of departments follow to decide the best matching. It summarizes the long steps of reviewing the contents of the course to be taught and the profiles of faculty members. In addition, the proposed solution solved the problem of time-consumption, which is caused by manually performing this job on a huge amount of data, and produced more accurate results.

The decision-making process for course distribution is made more challenging by the rise in the number of Ph.D. holders working in HEIs and the diversity of their research interests. Therefore, more values are required to improve this process.

The author in [3] has highlighted that the proper use of the information that is accessible across institutional repositories and the definition of what information may be shared are the first steps in resolving challenges in HE.

2.2. Educational Ontology

Traditionally, data are presented in semi-structured formats such as tabular representation, spreadsheets, and Web databases. These data types, unlike relational databases, are simple structures that are not in schema form. Humans can easily understand this unformal representation of data while machines cannot, since they are not framed in a specific schema [4]. Data are made available in digital form through ontologies. Thus, they are prepared to be shared and utilized to create knowledge-based systems for both humans and machines [5].

The nature of educational data might provide beneficial possibilities for the educational institutions if represented semantically to enhance their performance, making the usage of

Semantic Web (SW) technologies in education crucial [6]. Due to the ability of educational ontology to solve major problems such as knowledge modeling and information overload, it could be essential to employ it to solve many challenges in the education domain. SW has been extensively engaged in many studies within the education field over the past 10 years. These studies have played a key role in resolving some of the most challenging problems in various fields, including information integration and sharing, Web service annotation and discovery, and knowledge representation and reasoning [7].

2.3. Linked Data

Several governmental organizations have produced a large amount of data in the previous decade, leading to active research in various data engineering disciplines such as data representation, storage, and access. One of these research areas is linked data (LD).

The benefits of using this technique are:

1. *Uniformity*: Linked data are published in the form of a Resource Description Framework (RDF). This representation is expressed as triples that consist of subject, predicate, and object. All the triple components are defined as Uniform Resource Identifiers (URIs) that make each a unique identifier.
2. *Dereferenceability*: Each URI can be used to retrieve and locate information on the Web.
3. *Coherence*: URIs as triples can be used to establish a link between two datasets via the URI that represents a subject in a source dataset and the URI that represents an object in another (target) dataset.
4. *Integrability*.
5. *Timeliness*: Publishing and updating LD is straightforward, since it does not require loading and transforming.

Tim Berners-Lee proposed a five-star scheme [8] for encouraging individuals to publish in a linked open data (LOD) environment in 2010:

1. Open data that are available on the Web with an open license;
2. As structured data that can be read by machines;
3. As an open format;
4. All of the above, plus use W3C open standards to identify things (RDF and SPARQL), so that people may point to your content;
5. All of the above, plus link your data to the data of others to add context.

With the expansion of SW technologies, many research centers, institutions, and enterprises are publishing their data on the Web as LOD. Due to the spread of this technique's usage around the world, there was a need to create a global data cloud, and this was the main idea behind inventing the LOD. The LOD cloud began with 12 datasets in 2007. As of May 2020, this network contained 16,283 links from 1301 produced open datasets from different domains, such as government, companies, media, life science, publications, social media, scholarly data, etc. On the other hand, this gives a chance to third parties to take advantage of these open data to expand their information. Large institutions such as universities and HEIs compete to use these revolutions to improve their information systems.

Although many LD researchers face challenges in using this technique, benefits cannot be ignored, such as transparency, reusability, knowledge discovery, and interoperability [9] for different application areas.

LOD is the result of releasing LD under open licenses [10], which increases data reuse [11]. Integrated data often aids in the formation of comprehensive knowledge, which in turn supports decision-making. In addition, LD can answer complex queries that single datasets cannot answer, by using combined data from different sources.

One of the particular strengths of the LD approach is that it accepts heterogeneity and provides interoperability based on links between different datasets [12].

The goal of LD is to provide machine-readable connectivity between various data sources on the Internet. As a result, LD has been regarded as one of the most successful components in resolving many issues that educational institutions confront throughout decision-

making processes. Collecting content from resources, looking for missing academic information, and so on can be made easier and more precise, ensuring quality in HEIs.

Within universities and research centers, academic-teacher-related judgments are frequently influenced by their publication data first and foremost. Evaluating the content of academic teachers rules a variety of decisions within universities. That includes the positions the academic teacher can take, the courses they can teach, the projects they can be involved in, the training courses they require, and many more. In this research, these goals are readily accomplished by finding the proper scholarly open data and linking it to the local data of universities.

This work uses the local data of the Faculty of Computing and Information Technology (FCIT) within King Abdulaziz University (KAU), which is represented semantically to propose applying the LD technique to automatically enrich the local data and support the decision-making process for assigning courses to the most proper academic references.

3. Related Work

LD is one of the most powerful frameworks in the data management field; so, there is a significant presence of this golden research subject in different domains. Many researchers reported in their publications different approaches to automatically enrich and populate their ontology models.

Recently, LD and open data techniques seem very promising in HE and propose notable research in this area. Since 2009, LD has been established by educational domains to be used in many aspects to overcome many challenges [13].

One of the early tasks proposed in this domain that serves both students and academic teachers is leveraging LD to develop open universities. Many educational institutions are offering free open access to their educational resources to make online learning more widespread. On the other hand, they can find accurate information available as open educational data to enrich their data. Open universities in [14–16] are produced using source data of universities and external repositories of educational datasets. In [15], the researchers have applied some scenarios for the proposed architecture. Firstly, students need to check the related materials that support their decisions about their university and facilities' offered choices. Generating links between these choices and the opened educational materials and providing them in one dataset offers significant benefits for students. A student may become interested in certain topics or courses and will need to specialize in the supporting materials to supplement their knowledge with high-quality resources. On the other hand, the student could find some difficulties in studying some courses, which would make them change their decision. Secondly, the faculty member can have the chance to develop or renew the curriculum of the course they teach, after comparing it with the syllabus that is provided by the other linked universities.

The open university in [17] has described information about published materials, teacher research work, titles, courses, and audio-visual educational resources using semantic technology. By establishing a SPARQL endpoint, these data can be reused and made available to others. Since some universities have transitioned from traditional to digital learning by providing open educational resources (OERs), the LD vision exemplified by the software interface enables a new generation of OERs and open course ware (OCW) that can be semantically described and connected with other data and discoverable sources. These resources contain tools and materials that can be freely accessed, reused, modified, adapted, and shared in order to promote education. Linked open course ware data (LOCWD) is a language created by the researcher utilizing W3C's RDF technology. It uses the Internet to connect OERs, open licenses, OCW repositories, and other academic materials. The fundamental goal of these vocabularies is to link the stated OCW domain to LOD cloud datasets.

The study in [18] proposes a task-interaction framework for mobile learning to aid educational decision-making. The framework is built on the links between the various sorts of interactions that occur in a mobile learning activity and the pedagogically relevant tasks

for the activity. A case study has been created to show how the task-interaction framework might be applied to learning scenarios using mobile devices. The researchers have used MeLOD8, a mobile environment for learning with LOD, to apply the scenarios.

The researcher in [10] has examined the capability of LD and the sufficiency of the existing data source to promote student retention, progression, and completion. The researcher in this work used LD technology to develop an academic predictive model that targets first-year students at universities. They have applied two experiments. The first one predicts the students' likelihood of being at-risk. The second experiment uses easily accessible data from internal institutional data sources/repositories and external open data sources to forecast the academic performance/marks of the students. The sufficiency of LD and external opened data sources has been examined using questionnaires (surveys).

Under the fast growth of scholarly data, a significant number of studies have used LD to enrich the quality of the available researchers' data. In [19], a subset of scientific publications called CONICET Digital is published as LOD. The producers of this work have used the strength of SW and LD technologies to improve the recovery and reuse of data in the domain of scientific publications. Moreover, they considered the SW standards and reference RDF schemas such as Dublin Core, FOAF, and VoID. They convert and publish their data using the same guidelines for publishing government-linked data. On the Web of data, the data is linked with the external repositories DBLP, WIKIDATA, and DBpedia. The resulting platform particularly retrieves information from the scientific domain by combining data from different sources. Moreover, it allows users to view the resulting information related to the available data and run queries using the SPARQL language.

Ontura-Net [20] is a research project that employs LD approaches to explain the scientific activity of Ecuadorian university scholars. Under the realm of university scientific research activities, this study demonstrates the outcome of the LD manufacturing cycle. It is a legal term that refers to Ecuadorian university quality accreditation regulations. The main objective of this project is to assist universities in improving certain aspects, such as incorporating scattered teacher-researcher production into the network, which is crucial when establishing scientific and academic research information metrics from individuals or groups at the institutional level. It also aids in the identification and formation of scientific collaboration networks as well as the detection of priority potential domains in which legislators can assist in the formulation of science and technology policies.

Another Ecuadorian study [21] generated links between multiple bibliographic sources to find similar research areas and prospective collaboration networks through a combination of ontologies, vocabularies, and LD that enrich a base data model. The researchers linked diverse Ecuadorian HEIs with external scholarly data from bibliographic sources, such as Microsoft Academics, Google Scholar, DBLP, and Scopus, which make available their data via APIs. The resulting links are utilized to create a prototype that provides a centralized repository with bibliographic sources and allows academics throughout Ecuador to locate similar knowledge areas using data mining techniques.

The proposed work in [22] has solved the most common problems related to publications such as incomplete information, lack of semantic information, and author ambiguity, when two or more authors could share the same name or two or more names belong to one author. The external sources I-Scover and DBPedia datasets are utilized, considering the names in English and Japanese to deduplicate records and reduce data redundancy in publication data, extract more information about authors of articles, and tackle the problem of author ambiguity. The authors first normalize entity names before searching DBPedia for all available candidates. Then, they use semantic data from I-Scover and DBPedia to create semantic profiles for both entities and applicants. Finally, they use a combination of lexical and semantic profile similarities to find the equivalent DBPedia entity.

The researchers in [23] have developed a search engine called WISER. This system uses the benefits of the semantic approach and LD to find academic experts in the academic domain. It retrieves academic authors whose expertise is described through the publications they have produced, and it is relevant for a user query.

ScholarLens is an approach that is described in [24]. It aims to extract competencies from research publications using SW and LD techniques to generate user profiles automatically.

In [25], the study investigates the use of ontologies and LD to support the representation of researcher profiles in the academic environment. It describes an ontology model that is automatically populated. Bibliographic records are extracted from the DBLP repository to enrich the proposed ontology.

Based on the review of the related works, we can establish that the use of ontology and the LD technique has proven itself in the academic domain for different tasks. In addition, open-linked scholarly data were a solution for many problems related to publications, such as detecting similarities between authors' publications for scientific collaborations. On the other hand, we can state that no research from the related works finds the similarity between the academic staff publications and the topics of the taught courses and employs it to support the academic decision process, especially not to improve the decisions of course–teacher assignment.

4. Methodology

A massive amount of educational data is produced by different educational institutions every year [10]. These materials would be hard to discover or integrate into traditional information systems. That means, everything we need is available, but it is hard to find. As a result, wading into applying semantic and LD technologies in education would be crucial, since the nature of such educational data can generate opportunities for educational institutions to improve their performance and support the decision-making process.

Course–teacher assignment is one of the most common considerations that universities face regularly. It incorporates evaluating the academic teacher and determining their capacity to perform an assigned task, which traditionally passes through complicated processes, similar to many other educational decisions. Performing this task manually on this amount of data is inefficient, ambiguous, and time-consuming. Furthermore, some academic profiles have missing data or materials that are overburdened.

There is also a necessity to match the academic teachers' various qualifications to the course specifications. This step requires collecting more information from external sources.

As a proof of concept, this research uses King Abdulaziz University (KAU) data, with the Faculty of Computer and Information Technologies (FCIT) serving as the case study. KAU's staff committees and courses description are presented semantically in our previous work in [1]. This will be updated regarding most aspect elements that affect the course–teacher assignment decision. After that, external repositories will be searched to select the most appropriate dataset that enriches the needed information by generating link data.

Choosing the proper methodology for generating LD relies on different factors such as the case study or the scenario of the problem to be solved by this technique, the nature of the data, and the characteristics of the domain.

In the literature, few researchers have briefly described the methods and tools they use to operate in generating, linking, publishing, and using LD. One of the first studies [26], titled "A Cookbook for Publishing Linked Government Data on the Web", was published in 2011 and discussed the applied methodology. Most of the studies have followed the main steps mentioned in this book and can be summarized in the three following steps:

1. *Initialization*: This step includes specifying requirements and business objectives and then analyzing the datasets used in LD generation. Moreover, it involves selecting vocabularies and developing other specifications for metadata description.
2. *Innovation*: The process of combining datasets into a knowledge graph style. This includes data access, transformation, and enrichment.

    a.　For pilot applications, the developer needs to select the generic component and customize the needed tools, i.e., specifying LD components required in the domain of interest.

> b. Development of specific tools: implementing security measures to deal with the risk of communication.

3. *Validation*: The last phase, which is a continuous process. It comprises the reuse of open-source tools, improving components based on feedback, and testing data.

The LD in this study is created by employing the selected method and expanding it (as shown in Figure 1) in the following phases:
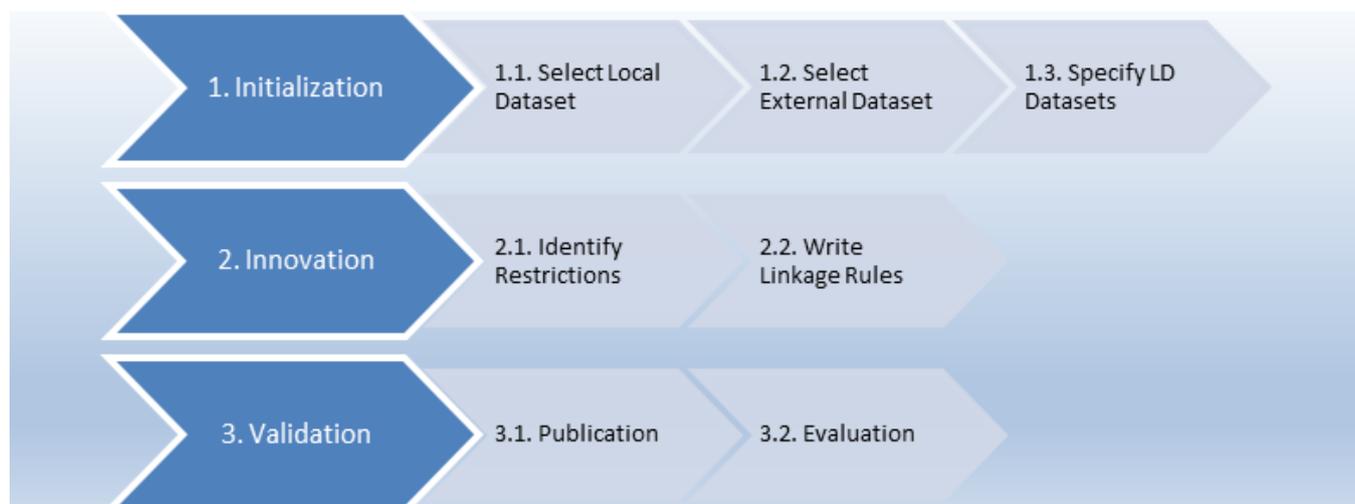
Figure 1. LD generation methodology.

4.1. Initialization

4.1.1. Select the Local Data Source: University Ontology

This step includes defining the requirements for selecting the proper data sources that include all possible elements that describe academic scholars and publications. First, the most effective elements in the academic decisions are specified. After that, the data sources are selected and accessed under the control of the following conditions:

- Availability and license;
- Representation in a machine-readable format such as Excel, CSV, XML, RDF, etc.;
- Covering the data that describes the scholarly activities of the selected academic teachers in the case study.

Like the first step, a survey is conducted to find the most elements that affect the course–teacher assignment decision. The results in Section 5.1 are employed to improve the research on three axes:

1. Select the local dataset and present it semantically;
2. Choose the most proper external dataset (scholarly data);
3. Query the resulting dataset to find the best course–teacher assignment.

The survey consists of four sections. The first section investigates the general information and measures, the experience of each department head. It is followed by two sections that examine how elements of both the courses and the academic reference influence the course–teacher assignment decision. The last section measures how students' feedback affects the decision. The main elements tested in the survey are described in Table 1.

Table 1. Survey elements description.

Element Type	Element Description
Course Elements	a. Course type (e.g., elective, required, etc.)
	b. Level/year at which the course is offered (e.g., 2xx, 3xx, 4xx, etc.)
	c. Course topic area (if the course matches the academic teacher's specialty), e.g., network, programming, algorithms, Algebra, etc.
	d. Course learning outcomes (CLOs): i.e., subtopics of the course that describe the knowledge, skills, and values that the course offers
	e. Course nature (theoretical or practical)
	f. Course teaching and learning strategies (e.g., e-learning, distance learning, group working projects, etc.)
	g. Course assessment methods (e.g., electronic exams)
Academic Profile Elements	a. Academic teacher's academic rank (professor, associated professor, assistant professor)
	b. Academic teacher's amount of teaching expertise (years)
	c. Academic teacher's preferences (preferred selection courses)
	d. Academic teacher's research area of interest during the last three years
	e. Academic teacher's amount of research in that area during the last three years
	f. Academic teacher's training courses that relate to the course's topic area and have been achieved during the last three years
	g. If the academic teacher holds an accredited international professional certificate to teach the course (e.g., Microsoft, ACCA, CPPS, etc.)
	h. If the academic teacher is a coordinator of the course
	i. If the academic teacher has coordinated a pre-requisite or co-requisite course (if any)
	j. If the academic teacher has taught the course before
	k. If the academic teacher has taught a pre-requisite or co-requisite course of the course (if any)
Students' Feedback (if the academic teacher has taught the course before)	a. Students' achievement in that course (i.e., students' results and distribution of grades)
	b. Students' satisfaction rate for the course quality evaluation
	c. Students' satisfaction rate for the academic teacher evaluation

As mentioned previously, this research uses the dataset of KAU, particularly the courses' details and members' profiles of the three departments of the FCIT. Since this targeted academic data was not available in RDF format, it was presented semantically in our previous work [1], based on the accreditation categorization of HE in Saudi Arabia.

The use of SW technology to support the decision-making process within universities was proposed in our previous research [2]. The ontology is called KAUONT, and it is created using Protégé. We have queried the data using SPARQL queries under the rule of not publicly disclosing information about members and academic data within educational institutions. Therefore, the results were published as quantitative data instead of qualitative. Using LD to improve the results and automatically enrich the local data was one of the future works that have been mentioned in our previous research, and it is the main task of this paper.

Therefore, it is used to characterize the local data. In addition, it is improved, regarding the survey results, by adding some classes and properties. Results from both studies will be compared in Section 6 to prove the success of using the LD technique.

4.1.2. Select the External Data Source: Scholarly Data

Under the fast and continuous growth of the scientific literature that brings difficulties among the high volume of published papers that need annotations and management, the number of novel technological infrastructures are found to help researchers and research institutions to easily browse, analyze, and forecast scientific research. Therefore, well-known bibliographic repositories are available online to extract scientific publications data from, such as DBLP and DBpedia. Using semantic repositories that use ontology as semantic schemata increase the possibility of automated reasoning about the data and allow implementation, since the most essential relationships between concepts are incorporated into the ontology.

On the other hand, another innovation is found, known as scientific knowledge graphs, which concentrates on the bibliographic domain and consists of metadata that describe research publications such as authors, venues, affiliations, research areas, and citations. This type of data representation contains a large number of entities and relations that are usually structured as RDF triples. These structured representations can support different tasks such as question answering, summarization, and decision systems. Some examples of scientific knowledge graphs are Open Academic Graph, Scholarlydata.org, Microsoft Academic Graph (MAG), Scopus, Semantic Scholar, Aminer, Core, OpenCitations, and Dimensions.

To choose the most proper external dataset, several scholarly repositories and scientific knowledge graphs are reviewed:

1. *Databases and logic programming (DBLP)*: DBLP is a bibliography that specializes in the computer science area. It contains the metadata of publications, authors, journals, and conference proceedings series. The dataset is available in RDF format to be freely reused at (https://dblp.org/rdf/) (accessed on 20 May 2022). The size of the March 2022 release of the DBLP dataset is over 252,573,199 RDF triples and 6,010,605 publications written by 2,941,316 authors. The developers are planning to develop a SPARQL endpoint to query the dataset easily.

2. *DBpedia*: DBpedia has been developed by Leipzig University and the University of Mannheim to represent the data of Wikipedia semantically. It is considered the most famous used dataset in LD projects, since it covers data from different domains and includes links to other datasets. Altogether, the DBpedia 2016-04 release consists of 9.5 billion pieces of information (RDF triples), out of which 1.3 billion were extracted from the English edition of Wikipedia.

3. *OpenCitations*: OpenCitations is expended of the Open Citations Corpus (OCC), which is considered as a scholarly citation repository that is licensed as open data for download and reuse under the Creative Commons public domain dedication. It provides accurate RDF-formatted bibliographic data. These are built using the SPAR Ontologies, according to the OCC metadata model, and are freely available for anybody to build upon, enhance, and reuse for any purpose, without copyright other database law restrictions. The OCC is updated continuously with new information from the academic literature. The whole OCC can be queried via the SPARQL endpoint (http://opencitations.net/sparql) (accessed on 14 April 2022). In addition, a data dump is available to download on the website: http://opencitations.net/download (accessed on 16 April 2022). OCC contains 302,758 citing bibliographic resources and includes information about 12,830,347 citation links to 6,549,665 cited resources.

4. *Scopus*: Scopus is a publication database created by Elsevier, which extracts bibliographic information, including citation records. In addition, it contains linked data about papers published in refereed journals, proceedings of conferences, and scientific book chapters. Scopus is considered as a RESTful API that provides obtained publication data. Unfortunately, this service is not free of charge; only institutions with a Scopus license can access it.

5. *WikiCite*: WikiCite is a bibliographic database based on Wikidata and founded in 2019. It represents scholarly article metadata. WikiCite's goals include improving citations in Wikimedia projects and creating an open, collaborative archive of bibliographic

data for new applications. There is a special SPARQL endpoint available to query WikiCite data online.

6. *ArnetMiner:* ArnetMiner is an online service available for indexing, searching, and mining huge scholarly data. It was established as a research study in social analysis, social communication ranking, and network extraction. It automatically collects the researcher's details from the Web. It uses heuristic rules to extract the information from online digital libraries. The dataset links researchers, conferences, and publications from different institutions. The citation data is collected from many scholarly sources including DBLP and ACM. It is available online at: https://aminer.org/billboard/AMinerNetwork (accessed on 18 August 2022). It includes about 2 million articles, 2,092,356 papers, 1,712,433 researchers, and 8,024,869 citation relationships.

7. *Microsoft Academic Knowledge Graph (MAKG)*: MAKG is a huge scholarly dataset including information about the related objects to a scientific publication, such as authors, journals, institutions, and areas of study. It is the RDF version of the Microsoft Academic Graph dataset (MAG). It is available as text files with a subscription, and it is freely accessed under Open Data Commons Attribution License (ODC-By) v1.0. The SPARQL Endnote for this dataset is available at: https://makg.org/sparql. The dataset is also available as RDF dump files and can be accessed directly at: https://makg.org/rdf-dumps/ (formerly http://ma-graph.org/rdf-dumps/) (accessed on 12 June 2022). The size of this knowledge graph is over 8 billion triples: 209,792,741 publications, 1,380,196,397 references, and 151,355,324 authors.

After a thorough examination of the selected scholarly data sources, MAKG is chosen as the source for the data extraction, due to the huge size of researchers' data and the detailed structure of the dataset that is available on the MAKG website. In particular, the dataset offers the needed information about the authors, publications, and citations as well, and it is easy to query the dataset using the available SPARQL endpoint to select the authors from KAU and count them. To test MAKG, two queries were run, as follows:

1. Count the authors from KAU by finding the number of the authors in each affiliation, as shown in Figure 2.
2. Check the availability of all the needed information in the MAKG endpoint. To select all the data of the authors from KAU, a SPARQL query is run on the MAKG endpoint, as described in Figure 3.
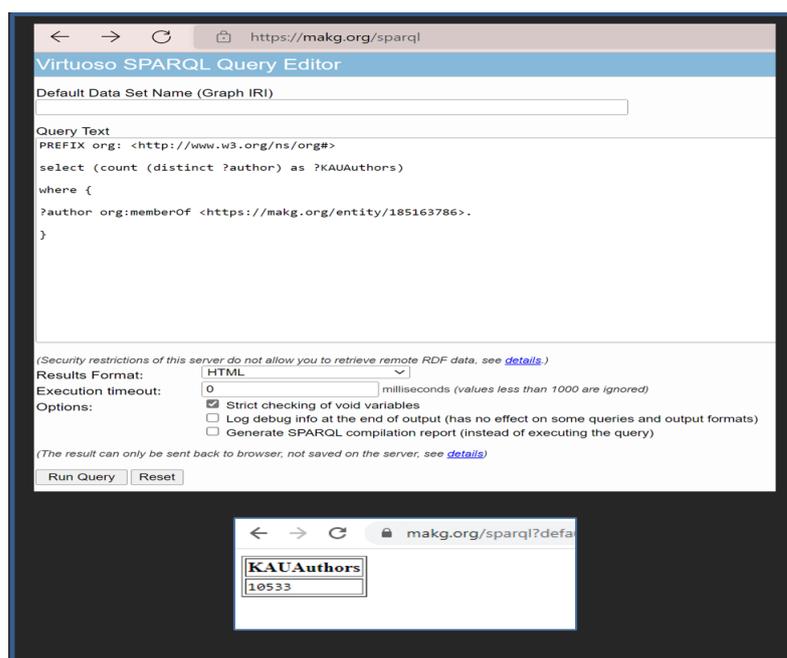
Figure 2. Count KAU researchers' profiles in MAKG.

Figure 3. Querying MAKG to list some KAU researchers' attributes.

4.1.3. Specify LD Dataset

This is the most sensitive step for link generation. It includes the following:

- Specify the access method to the datasets. Since the link is generated between an RDF file and an online knowledge graph, the KAU RDF local dataset (KAUONT) is loaded, and the SPARQL endpoint of MAKG is pointed using the silk workbench editor.
- Identify classes with instances that can be the subject of linking. The link is performed by connecting the two datasets by the academic staff name and the affiliation name.

4.2. Innovation

The link can be generated manually in the case of the small datasets, but, because this study is applied to a larger dataset, performing the manual link is not feasible. Silk [27] is the chosen tool in this research. It is an open-source tool that has a discovery engine that offers very significant features, as follows:

- A declarative, flexible language for defining linking rules;
- RDF link creation assistance such as owl:sameAs links;
- Work in a distributed environment through accessing local and remote SPARQL endpoints;
- Useful when terms from multiple vocabularies are combined and there are no standard RDFS or OWL schemata;
- High performance and scalability data management;
- Network load reduction by caching and reusing SPARQL result sets.

4.2.1. Identify Restrictions

This step limits the link to the target set (MAKG) of the external data and reduces the linkage time in Silk. Since the case study is applied to KAU academic staff, the restriction aimed to limit the link to the members of KAU only, as shown in Figure 4.

Figure 4. Using target restriction to limit the link.

4.2.2. Write Linkage Rules

To generate the link the following rules were applied to datasets:

- Specify how resources would be compared. The two main entities to be compared are kau:memberName from KAUONT with foaf:name from the MAKG dataset. Both will be transformed into lowercase to prevent any mismatch caused by the use of lower and upper cases.
- The output will be compared using the "Levenshtein distance" metric in the SILK workbench to guarantee the exact match of similarity (as shown in Figure 5).
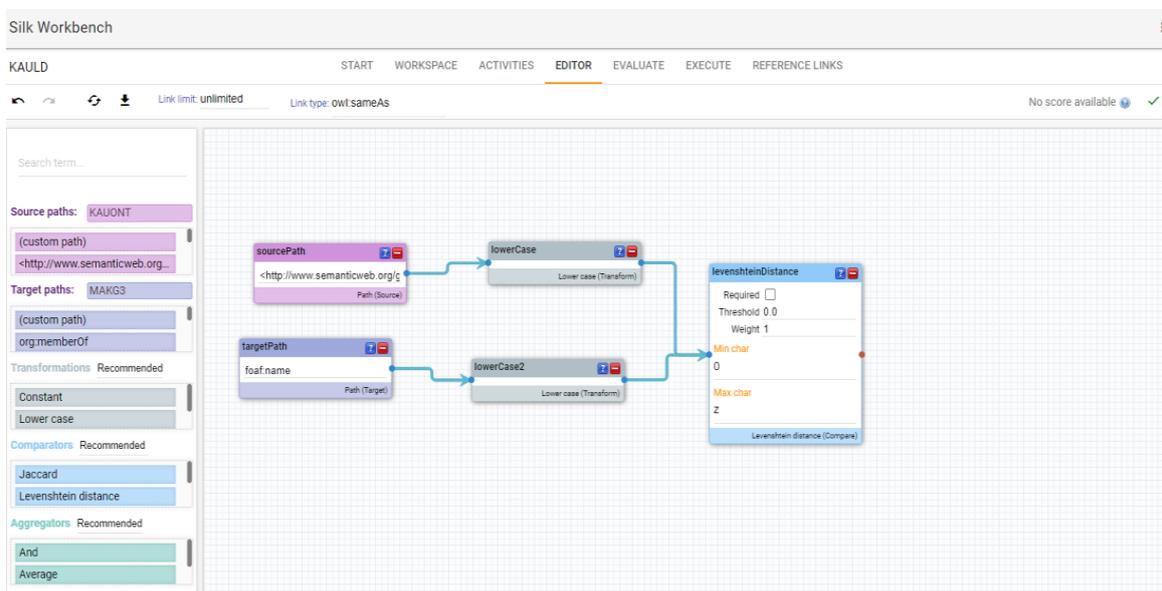
Figure 5. Generate the link using Silk Workbench editor.

The results, as shown in Figures 6 and 7, include 150 links.

Figure 6. Part of the link results.

Figure 7. Number of the resulting links, as shown in Silk.

4.3. Validation

Validation is the process that follows link generation and guarantees the effective use of the resulting link data. It consists of the following:

1. *Publication:* The resulting linked dataset (KAULD) is published to provide machine access to it, using tools such as GraphDB.
2. *Evaluation:* KAULD is evaluated to retrieve the new data. This task is performed using a federated SPARQL query that includes the 'SERVICE' statement, because of the need to send a query to a remote site (MAKG endpoint). The prefixes in Figure 8 are used every time KAULD is queried.

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX kau: <http://www.semanticweb.org/ghade_000/ontologies/2019/4/DSS2.owl#>
PREFIX dcterms: <http://purl.org/dc/terms/>
PREFIX fabio: <http://purl.org/spar/fabio/>
PREFIX prism: <http://prismstandard.org/namespaces/basic/2.0/>
PREFIX mag: <https://makg.org/property/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
```

Figure 8. Prefixes used when querying the resulting data.

5. Results and Discussion

5.1. Survey

The survey collected the response of 41 heads of departments at KAU. Most of the participants have been in their current position for at least 2 to 4 years, which indicates that most of them have a reasonable understanding of the system. When asked if they have faced problems with the current course-distribution method, 73.2% find that the process is complicated, since it is processed manually, while only 10% feel that the course assignment procedure is smooth. This response supports the main motivations of this research.

Regarding the course elements shown in Figure 9, the survey proved that course topics strongly affect the decision of assigning the teacher to teach a specific course, since more than 45% of the participants support it. Course type is also indicated as another important element, as about 30 participants strongly support it. On the other hand, the majority of the participants find that the other course elements are not significant, and they do not rely on them when producing their decisions.
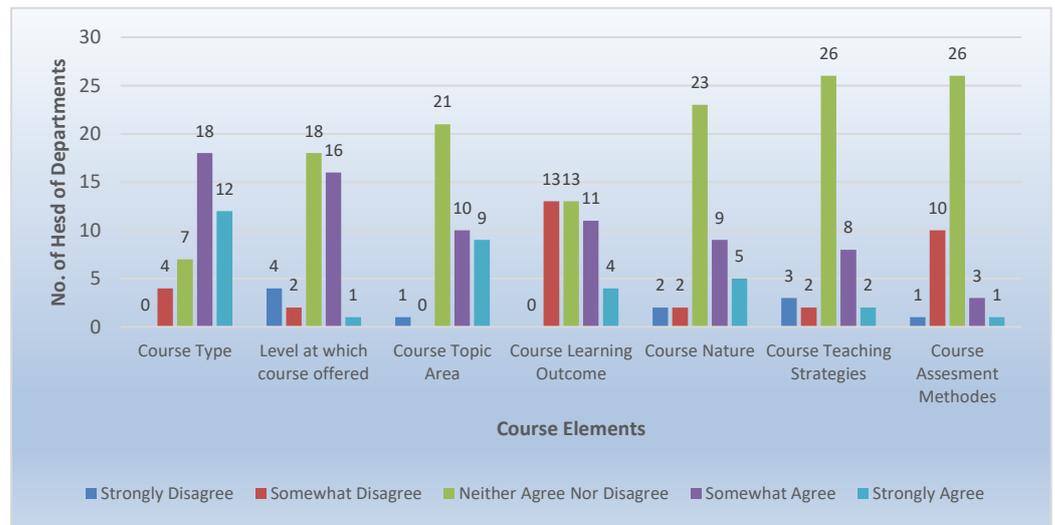
Figure 9. Comparing course elements.

According to the testing of the academic profile elements (summarized in Figure 10), the survey depicts the impact of a teacher having taught the course before. The majority of participants believe it is a significant factor in deciding course–teacher distribution. No participant gave a vote against it. In addition, it proves that the research area of the academic teacher is seriously a considerable element that controls the decision, as more than 60% of the votes strongly support it. The academic rank was considered by more than half of the participants. This element can be used to set the teachers' priority, when more than one teacher is allocated to teach a specific course. Furthermore, the survey indicated the importance of the certificates the academic teachers have in the course–teacher assignment. For a teacher to be the course coordinator, the result shows how essential it is, since 20% strongly believe, and 56% support it, since course coordination plays a significant role in course distribution, while only 12% of people disagree with it. Regarding these results, the affective teacher elements were teaching the course before, the research area of the academic staff, and coordinating the course.
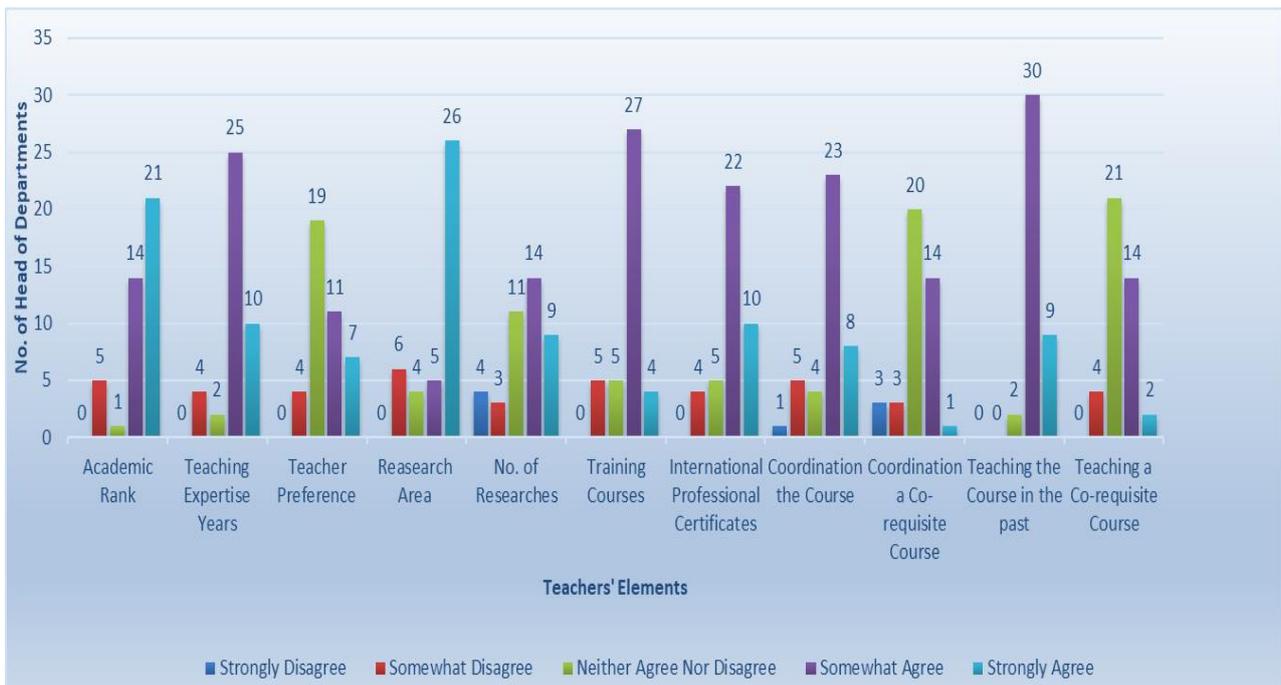
Figure 10. Comparing academic teacher elements.

On the other hand, Figure 11 proves that students' achievement and feedback are usually not considered when assigning teachers to courses.
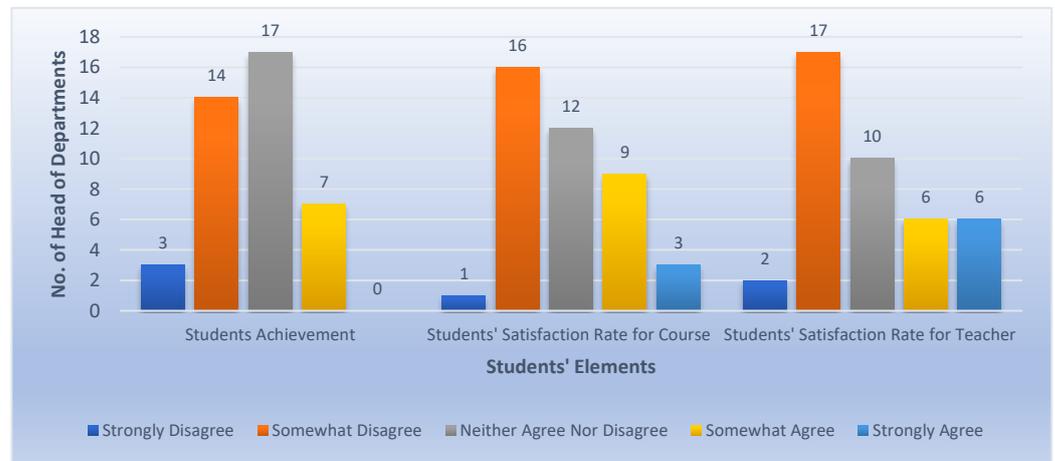
Figure 11. Comparing students' feedback elements.

5.2. Resulting Linked Data

To judge the success of using the link data technique in improving educational decisions, KAULD is tested using federated queries to select all the academic teachers who can teach courses from the same department. The selection relies on the factors from the survey result mentioned in Section 5.1 and the elements extracted from MAKG, as shown in Figure 12.
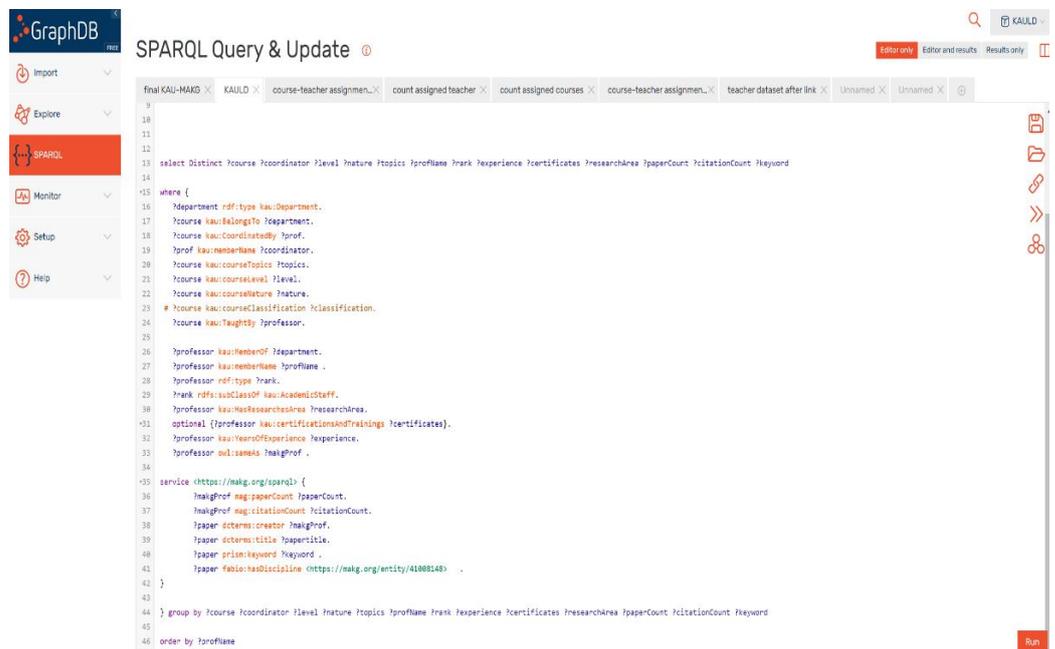
Figure 12. Example of extracting data from KAULD using federated query.

Table 2 summarizes the quantitative data from each department's courses and academic staff. The evaluation was limited to Ph.D. holders, because they are typically involved in research and publish in journals and conferences. Furthermore, they teach the courses, while the majority of non-Ph.D. holders are deemed teaching assistants. As a result, it is rare to find scholarly records of non-Ph.D. holders. Simultaneously, all academic instructor profiles in the faculty, from all degrees, were translated into RDF format for future processing.

Table 2. Number of courses and academic staff in FCIT.

Department	Number of Courses	Number of Academic Teachers	Number of Ph.D. Holders Only	Number of Linked Academic Teachers
Computer Science	44	109	51	51
Information System	41	99	45	52
Information Technology	34	106	53	47

Our previous work [2] solved the problem of the course–teacher assignment by developing an educational ontology that models the semantics of the courses and academic profiles in universities. The results depend on two factors only: who taught the course before and the match between the course topic and the research interest of the academic teacher. Figures 13 and 14 summarize the results of course–teacher assignment using SW techniques only from the previous study [2]. It is shown in Figure 13 that a significant number of courses are not included. Around 7% only of courses in the Computer Science department are assigned to qualified teachers, while 29% of courses in the Information System department and 62% of courses in the Information Technology department are included in the course–teacher assignment results. Figure 14 shows that 40% of Ph.D. holders are assigned to teach courses in the Computer Science department, while 43% of Ph.D. holders in the Information System department and 58% in the Information Technology department are assigned to teach courses from the same department.
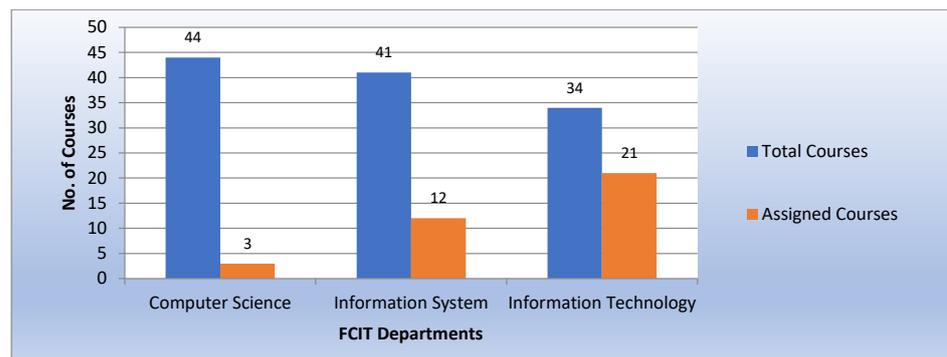
Figure 13. Assigned courses from the previous study [2].
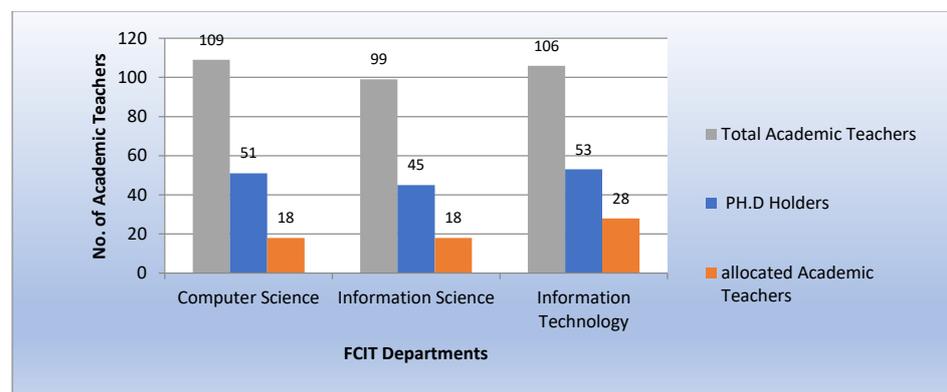
Figure 14. Allocate teachers from the previous study [2].

The query in Figure 15 counts the numbers of courses assigned to the most proper teachers in the possible matching, between the courses and academic teachers in this study. The query retrieves the courses from the KAULD dataset regarding the most effective elements mentioned in Section 5.1.
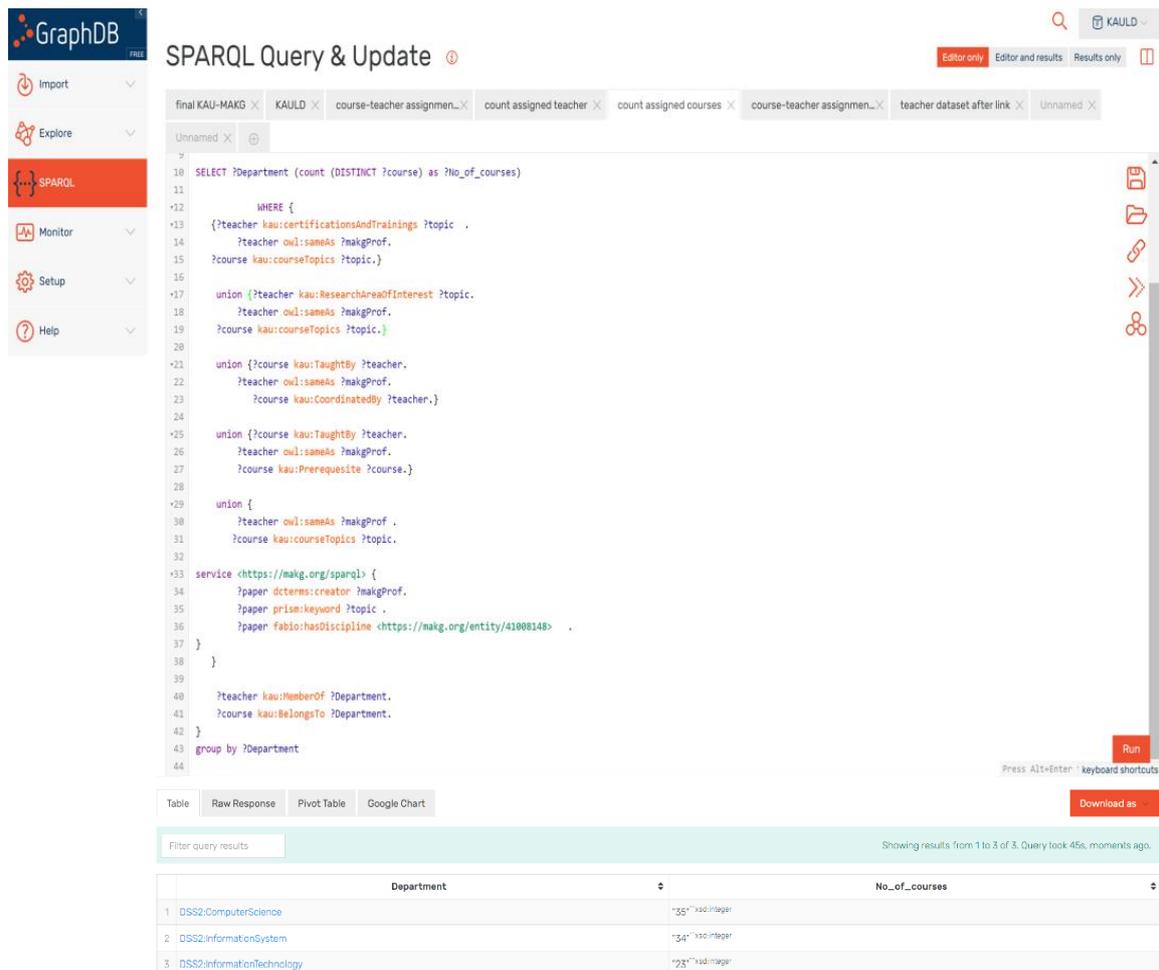
Figure 15. Number of courses assigned to teachers using LD.

As illustrated in Figure 16, more than 81% of Computer Science courses are assigned to teachers, with less than 20% remaining unassigned. It is also stated that teachers are assigned to around 83% of the courses in the Information System department. On the other hand, teachers in the Information Technology department are assigned to 68% of the courses. This result demonstrates that most of the courses matched the best academic references regarding their elements.
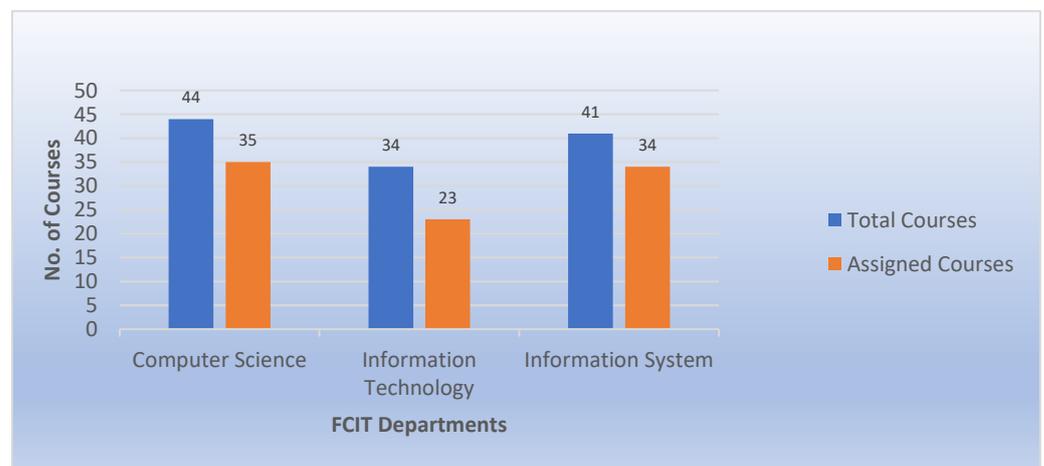
Figure 16. Comparing number of courses assigned to teachers using LD to the total number of courses.

Compared to that previous analysis from Figure 13, it is found that a larger number of courses are assigned to more qualified teachers, after enhancing the academic teachers' profiles using the LD technique, as more factors that affect the decision are considered in the query.

The query in Figure 17 counts the number of academic teachers from each department that are qualified to teach courses related to their qualifications in this study. The query matches the teachers with the related courses from the KAULD dataset, regarding the most effective elements mentioned in Section 5.1.
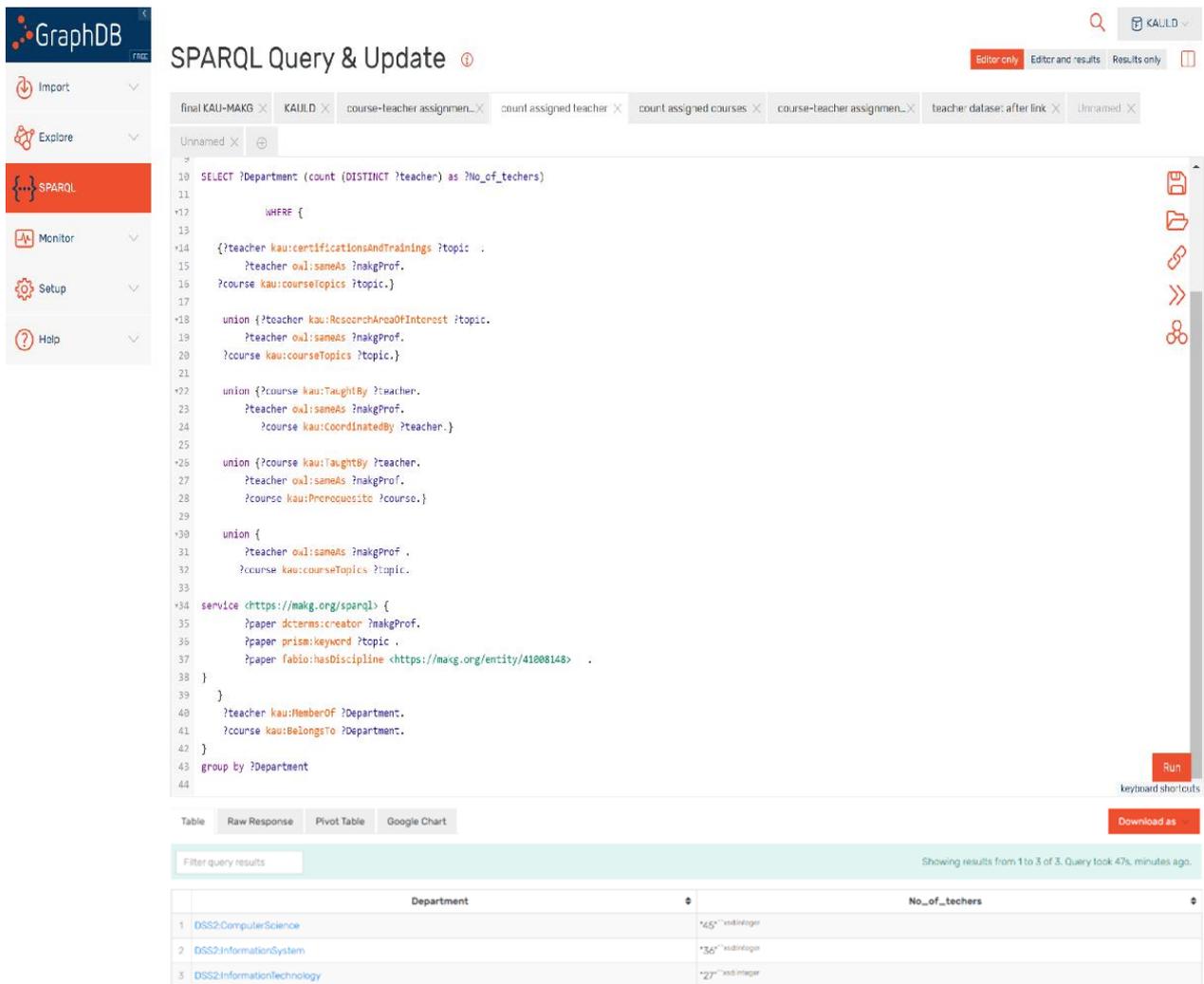
Figure 17. Number of allocated teachers using LD.

Figure 18 shows that approximately 88% of the academic teachers in the Computer Science department were assigned to teach courses, compared to roughly 70% in the Information Systems department. On the other hand, 68% of the academic staff in the Information Technology department are assigned to teach courses from the same department. Compared to the results from the previous study from Figure 14, most of the academic teachers in each department are assigned to teach courses that match their qualifications.
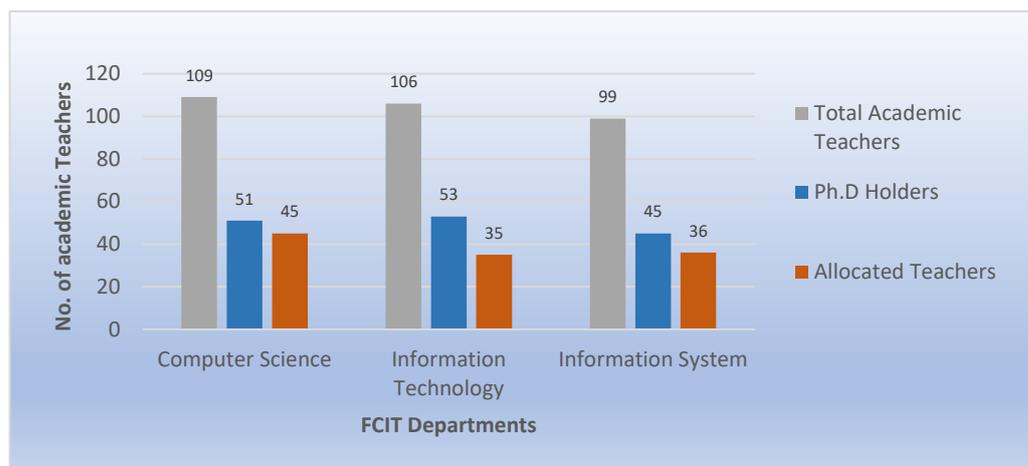
Figure 18. Comparing number of allocated teachers using LD to the total number of teachers.

To summarize the results, leveraging LD with SW techniques has succeeded in giving sufficiently accurate decisions. This proves that LD adds value to SW when employed to solve decision-making challenges within HE. Although the percentages that are shown in the evaluation process cover the most teachers and courses, there is still a need to solve some shortages such as allocating the rest of the teachers, setting priorities to organize choosing the most proper reference to teach the course in the case of locating more than one teacher, and assigning teachers to new courses. In the future, techniques such as machine learning and data mining can be applied to the resulting dataset to solve these issues.

6. Conclusions

Currently, many HEIs are modernizing decision support processes. This step leads to trendy research subjects. Therefore, several researchers have examined different techniques to solve the challenges caused by this modernization. LD is one of the most successful technologies, proposed by significant amounts of the related literature, solving many challenges in HE. The academic teacher is the crux of most decisions in HEIs. Since ranking academic teachers relies on their academic and research experience, there is a need to find a solution to enrich the academic teachers' profiles, especially their research records. This work enhances the decision-making process within universities by generating a link between a university ontology that represents courses and academic profiles semantically and an open scholarly dataset. Engaging LD technology enhances university data with needed or missed researchers' data related to their research activities, such as publications and citations.

The study aims to improve the previous results of mapping the most qualified academic teacher with a new course to teach. A survey was conducted to find the most effective elements that control this process. This experiment is applied using the Silk tool to generate the link between the semantic data of the Faculty of Computing and Information Technology with its three departments at KAU and the scientific knowledge graph MAKG. KAULD is the resulting dataset, and it was published using GraphDB.

A statistical analysis of the results was performed and compared to the results from the previous work. The comparison showed that LD succeeded in improving the decision-making process and, unlike using SW alone, the results of leveraging LD with SW included the majority of the courses and teachers. Most of the courses in each department are assigned to more qualified teachers from the same department, while teachers are allocated to teach courses most related to their qualifications.

Although most teachers have matched with most courses, several shortages have appeared, especially when providing new courses in a department or when more than one teacher is assigned to teach one course simultaneously. As a suggestion to solve these shortages, more artificial intelligence technologies, such as machine learning and data

mining, can be applied in our future work on the resulting dataset, to predict more course–teacher assignments and set teachers' priorities. In addition, the system can be extended to support more decisions within universities or to solve more educational challenges. Other universities can reuse it, especially those that apply the same rules as KAU.

References

1. Ashour, G.; Al-Dubai, A.; Romdhani, I.; Aljohani, N. An Ontological Model for Courses and Academic Profiles Representation: A case study of King Abdulaziz University. In Proceedings of the 2020 International Conference Engineering Technologies and Computer Science (EnT), Moscow, Russia, 24–27 June 2020; IEEE: New York, NY, USA; pp. 123–126. [CrossRef]
2. Ashour, G.; Al-Dubai, A.; Romdhani, I. Ontology-based Course Teacher Assignment within Universities. *Int. J. Adv. Comput. Sci. Appl.* **2020**, *11*, 720–728. [CrossRef]
3. Daniel, B. *Big Data and Analytics in Higher Education: Opportunities and Challenges*; Wiley: Hoboken, NJ, USA, 2015. [CrossRef]
4. Ristoski, P.; Paulheim, H. *Semantic Web in Data Mining and Knowledge Discovery: A Comprehensive Survey*; Web Semantics: Science, Services and Agents on the World Wide Web; Elsevier: Amsterdam, The Netherlands, 2016; Volume 36, pp. 1–22. [CrossRef]
5. Kabir, S.; Ripon, S.; Rahman, M.; Rahman, T. *Knowledge-Based Data Mining Using Semantic Web*; IERI Procedia; Elsevier: Amsterdam, The Netherlands, 2014; Volume 7, pp. 113–119. [CrossRef]
6. Tapia-Leon, M.; Rivera, A.C.; Chicaiza, J.; Luján-Mora, S. Application of ontologies in higher education: A systematic mapping study. In Proceedings of the 2018 IEEE Global Engineering Education Conference (EDUCON), Santa Cruz de Tenerife, Spain, 17–20 April 2018; IEEE: New York, NY, USA, 2018; pp. 1344–1353. [CrossRef]
7. Blomqvist, E. *The Use of Semantic Web Technologies for Decision Support—A Survey*; ISO Press: Amsterdam, The Netherlands, 2014. [CrossRef]
8. Bizer, C.; Heath, T.; Berners-Lee, T. Linked Data: The Story so Far. *Int. J. Semant. Web Inf. Syst. (IJSWIS)* **2009**, *5*, 1–22. [CrossRef]
9. Bernstein, A.; Hendler, J.; Noy, N. *A New Look at the Semantic Web*; ACM DL: New York, NY, USA, 2016. [CrossRef]
10. Sarker, F. Linked Data Technologies to Support Higher Education Challenges: Student Retention, Progression and Completion. 2014. Available online: https://eprints.soton.ac.uk/374317/ (accessed on 18 August 2022).
11. Alban, M.; Mauricio, D. Predicting University Dropout through Data Mining: A Systematic Literature. *Indian J. Sci. Technol.* **2019**, *12*, 4. [CrossRef]
12. Jabbar, S.; Ullah, F.; Khalid, S.; Khan, M.; Han, K. Semantic Interoperability in Heterogeneous IoT Infrastructure for Healthcare. Wireless Communications and Mobile Computing, Hindawi. 2017. Available online: https://www.hindawi.com/journals/wcmc/2017/9731806/ (accessed on 31 August 2022).
13. Pereira, C.K.; Siqueira, S.W.M.; Nunes, B.P.; Dietze, S. Linked data in Education: A survey and a synthesis of actual research and future challenges. *IEEE Trans. Learn. Technol.* **2018**, *11*, 400–412. [CrossRef]
14. Nahhas, S.; Bamasag, O.; Khemakhem, M.; Bajnaid, N. Linked Data Approach to Mutually Enrich Traditional Education Resources with Global Open Education. In Proceedings of the 2018 1st International Conference on Computer Applications & Information Security (IC-CAIS), Riyadh, Saudi Arabia, 4–6 April 2018; pp. 1–8. [CrossRef]
15. D'aquin, M. Putting Linked Data to Use in a Large Higher-Education Organisation. In *ILD@ ESWC*; Citeseer: Princeton, NJ, USA, 2012; pp. 9–21.
16. Zablith, F.; Fernandez, M.; Rowe, M. *Production and Consumption of University Linked Data*; Taylor & Francis (Routledge): New York, NY, USA, 2012. [CrossRef]

17. Piedra, N.; Tovar, E.; Colomo-Palacios, R.; López, J.; Chicaiza, J. *Consuming and Producing Linked Open Data: The Case of OpenCourseWare*; Emerald Group Publishing Limited: West Yorkshire, UK, 2014. [CrossRef]

18. Fulantelli, G.; Taibi, D.; Arrigo, M. *A Framework to Support Educational Decision Making in Mobile Learning, Computers in Human Behavior*; Elsevier: Amsterdam, The Netherlands, 2015; Volume 47, pp. 50–59. [CrossRef]

19. Zarate, M.; Buckle, C.; Mazzanti, R.; Samec, G. Improving Open Science Using Linked Open Data: CONICET Digital Use Case. *J. Comput. Sci. Technol.* **2019**, *19*, e05. [CrossRef]

20. Cadme, E.; Piedra, N. Producing linked open data to describe scientific activity from researchers of Ecuadorian universities. In Proceedings of the 2017 IEEE 37th Central America and Panama Convention (CONCAPAN XXXVII), Managua, Nicaragua, 15–17 November 2017; pp. 1–6. [CrossRef]

21. Sumba, X.; Sumba, F.; Tello, A.; Baculima, F.; Espinoza, M.; Saquicela, V. *Detecting Similar Areas of Knowledge Using Semantic and Data Mining Technologies*; Electronic Notes in Theoretical Computer Science; Elsevier: Amsterdam, The Netherlands, 2016; Volume 329, pp. 149–167.

22. Miao, Q.; Meng, Y.; Fang, L.; Nishino, F.; Igata, N. Link scientific publications using linked data. In Proceedings of the 2015 IEEE 9th International Conference on Semantic Computing (IEEE ICSC 2015), Anaheim, CA, USA, 9–11 February 2015; pp. 268–271. [CrossRef]

23. Cifariello, P.; Ferragina, P.; Ponza, M. *Wiser: A Semantic Approach for Expert Finding in Academia Based on Entity Linking, Information Systems*; Elsevier: Amsterdam, The Netherlands, 2019; Volume 82, pp. 1–16. [CrossRef]

24. Sateli, B.; Löffler, F.; König-Ries, B.; Witte, R. ScholarLens: Extracting competences from research publications for the automatic generation of semantic user profiles. *PeerJ Comput. Sci.* **2017**, *3*, e121. [CrossRef]

25. Bravo, M.; Reyes-Ortiz, J.A.; Cruz, I. Researcher profile ontology for academic environment. In *Science and Information Conference*; Springer: Cham, Switzerland, 2019; pp. 799–817. [CrossRef]

26. Hyland, B.; Wood, D. The Joy of Data—A Cookbook for Publishing Linked Government Data on the Web. In *Linking Government Data*; Wood, D., Ed.; Springer: New York, NY, USA, 2011; pp. 3–26. [CrossRef]

27. Silk The Linked Data Integration Framework. Available online: http://silkframework.org/ (accessed on 13 June 2022).