The Scientific Journal of Cihan University – Sulaimaniya **PP: 271-291**
Volume (7), Issue (2), December 2023
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

Adaptation of Academic Knowledge Representation from Heterogeneous Educational Data Sources

Lawchak Fadhil Khalid[1], Nawzat Sadiq Ahmed[2]

[1,2] Information Technology Department, Duhok Polytechnic University, Duhok, Iraq

Email: lawchak.fadhil@gmail.com[1], nawzat.ahmed@dpu.edu.krd[2]

Abstract:

Data in the modern era is diverse and always changeable. As a result of the accumulation of data from diverse systems has developed into a major issue or a hot topic in the process of adapting knowledge bases from various systems (i.e., Medical, Architectural and Educational systems, etc.). Furthermore, there are many academics who attempted to present a unified perspective of the information that is kept within any organization's relational, Extended Markup Language (XML), and Lightweight Directory Access Protocol (LDAP) data sources. Because of this, innovative approaches are required in order to collect, store, and analyze data from a wide variety of data sources in the most effective way that is feasible. Also, if the majority of data sources have already been included into the system that manages the data warehouse, there is still a demand for automated solutions that can generate a knowledge base for those responsible for making decisions in such fields. The primary objective of this research is to suggest a solution namely adapt academic knowledge representation from diverse educational data sources within a university and/or among other universities. This is done with the intention of sharing and making it easier for academic staff and students to gain access to knowledge that is helpful to them. The proposed system is developed using the V-model software development methodology, and research methodology used was mixed-methodology method. The proposed system was evaluated and proved to be a robust system that helped immensely with the tasks of Quality Assurance (QA) departments and also helped decision makers for universities to track the performance of their establishments.

Keywords: Knowledge Representation, Decision Support Systems, Data Collection and Processing.

الملخص:

البيانات في العصر الحديث متنوعة وقابلة للتغيير دائمًا. ونتيجة لتراكم البيانات من الأنظمة المتنوعة فقد تطورت إلى قضية رئيسية أو موضوع ساخن في عملية تكييف قواعد المعرفة من الأنظمة المختلفة (أي الأنظمة الطبية والمعمارية والتعليمية، وما إلى ذلك). علاوة على ذلك، هناك العديد من الأكاديميين الذين حاولوا تقديم منظور موحد للمعلومات التي يتم الاحتفاظ بها ضمن مصادر البيانات العلائقية و Extended Markup Language (XML) و Lightweight Directory Access Protocol (LDAP) لأي مؤسسة. ولهذا السبب، هناك حاجة إلى أساليب مبتكرة لجمع البيانات وتخزينها وتحليلها من مجموعة واسعة من مصادر البيانات بأكثر الطرق فعالية الممكنة. أيضًا، إذا كانت غالبية مصادر البيانات قد تم تضمينها بالفعل في النظام الذي يدير مستودع البيانات، فلا يزال هناك طلب على الحلول الآلية التي يمكنها إنشاء قاعدة معرفية للمسؤولين عن اتخاذ القرارات في مثل هذه المجالات. الهدف الأساسي من هذا البحث هو اقتراح حل وهو تكييف تمثيل المعرفة الأكاديمية من مصادر البيانات التعليمية المتنوعة داخل الجامعة و/أو بين الجامعات الأخرى. ويتم ذلك بهدف المشاركة وتسهيل وصول أعضاء هيئة التدريس والطلاب إلى المعرفة المفيدة لهم. تم تطوير النظام المقترح باستخدام منهجية تطوير البرمجيات V-model، وكانت منهجية البحث المستخدمة هي طريقة البحث Mixed-method. تم تقييم النظام المقترح وأثبت أنه نظام قوي ساعد بشكل كبير في مهام أقسام Quality Assurance (QA) كما ساعد صناع القرار في الجامعات على تتبع أداء مؤسساتهم. إن غياب نظام موحد لإدارة البيانات في المؤسسات الأكاديمية داخل

DOI: http://dx.doi.org/10.25098/7.2.34

إقليم كردستان، العراق، لا يحد من قدرتها على الاستفادة من البيانات في صنع القرار الاستراتيجي والابتكار فحسب.

الكلمات المفتاحية: تمثيل المعرفة، أنظمة دعم القرار، جمع البيانات ومعالجتها.

پوختە:

داتاکان لە سەردەمی مۆدێرن دا هەمەچەشنن و هەمیشە دەگۆڕدرێن. لە ئەنجامی کەڵەکەبوونی داتا لە سیستەمی جۆراوجۆرەوە پەرەی سەندووە و بووەتە پرسێکی سەرەکی یان بابەتێکی گەرم لە پرۆسەی گونجاندنی بنەماکانی زانیاری لە سیستەمی جۆراوجۆرەوە (واتە سیستەمی پزیشکی، تەلارسازی و پەروەردەیی و هتد). جگە لەوەش، زۆرێک لە ئەکادیمییەکان هەن کە هەوڵیان داوە دیدێکی یەکگرتوو لەو زانیاریانە بخەنەڕوو کە لەناو سەرچاوەی داتا پەیوەندیدارەکان، Extended Markup Language (XML) و پرۆتۆکۆڵی (LDAP)Lightweight Data Access Protocol ی هەر ڕێکخراوێکدا دەهێڵرێنەوە. بەهۆی ئەمەوە، ڕێبازی داهێنەرانە پێویستە بۆ کۆکردنەوە و هەڵگرتن و شیکردنەوەی داتا لە سەرچاوەی داتا جۆراوجۆرەکانەوە بە کاریگەرترین شێوە کە دەتوانرێت ئەنجام بدرێت. هەروەها ئەگەر زۆربەی سەرچاوەکانی داتا پێشتر خراونەتە ناو ئەو سیستەمەی کە کۆگای داتا بەڕێوەدەبات، ئەوا هێشتا داواکاری لەسەر چارەسەری ئۆتۆماتیکی هەیە کە بتوانێت بنکەیەکی زانیاری بۆ ئەو کەسانە دروست بکات کە بەرپرسیارن لە بڕیاردان لەم جۆرە بوارانەدا. ئامانجی سەرەکی ئەم توێژینەوەیە پێشنیارکردنی چارەسەرێکە کە بریتییە لە گونجاندنی نوێنەرایەتی زانیاری ئەکادیمی لە سەرچاوەی داتا پەروەردەییە جۆراوجۆرەکانەوە لە ناو زانکۆیەک و/یان لە نێوان زانکۆکانی تردا. ئەمەش بە مەبەستی هاوبەشکردن و ئاسانکردنی دەستڕاگەیشتن بە زانیاری بۆ ستافی ئەکادیمی و خوێندکاران ئەنجام دەدرێت کە یارمەتیدەرە بۆیان. سیستەمی پێشنیار کراو بە بەکارهێنانی میتۆدۆلۆژیای پەرەپێدانی نەرمەکاڵای V-model پەرەی پێدراوە و شێوازی توێژینەوەی بەکارهێنراو شێوازی Mixed-Method بووە. سیستەمی پێشنیارکراو هەڵسەنگاندنی بۆ کرا و سەلماندی کە سیستەمێکی بەهێزە کە یارمەتیدەرێکی بێئەندازە بووە لە ئەرکەکانی بەشەکانی Quality Assurance (QA) و هەروەها یارمەتی بڕیاردەران دەدات بۆ زانکۆکان بۆ بەدواداچوون بۆ ئەدای دامەزراوەکانیان.

کلیلە وشە: نوێنەرایەتی زانیاری، سیستەمی پشتگیری بڕیار، کۆکردنەوە و پرۆسێسکردنی زانیاری.

1. INTRODUCTION

Knowledge Representation Systems (KRSs) are one of the pillars of the advancements of any fields that are actively being developed and are encouraged to become a better system providing a better decision-making process in order to enhance such fields. This makes it one of the most important mechanisms to showcase how a certain institute is doing and how well their progress is going. The evolution of heterogeneous integrated data sources has become a topical issue as data nowadays is very diverse and dynamic. Hence, novel methods are necessary to collect, store and analyze data from various data sources as efficiently as possible, while also handling changes in data structures that have occurred as a result of evolution [1]. Dealing with heterogeneous data sources and integrating them has been focused on recently in many different fields such as Traffic Control [2], Healthcare Management [3], Supply Chain Management [4] and many more fields. Integrating data from multiple heterogeneous sources need to deal with different data models, database schema, and query languages as modern applications often need to manage and analyze widely diverse datasets that span multiple data models [5]. Nevertheless, our proposed system is a contribution for higher educational institutes that focuses on displaying educational performances for universities and institutes in the Kurdistan Region of Iraq (KRI). Apart from that, it is very supportive for learners and lecturers. This research aims to create a system that could gather information about academic activities of a specific university from multiple different sources including both digital files (such as PDF and Excel and CSV files) and also physical sources such as reports from QA staff, or sheets with academic activities listed, format them in a specific way and present them as knowledge base. From that it should help

DOI: http://dx.doi.org/10.25098/7.2.34

decision makers of educational institutions into making appropriate decisions for their students and to improve the educational process in their institutions. The proposed system would make decisions to measure performance of a specific university based on the numbers of academic activities per a certain list of categories that were marked by QA staff as important categories considered for university ranking in the NUR system.

The cost of implementing Extract Transform-Load (ETL) processes for data warehousing can be significant in some particular situations. The process of consolidating diverse data into a unified data model has the potential to negatively impact performance. Moreover, the curation of varied datasets and the maintenance of the pipeline can be a labor-intensive task. Consequently, there is a growing tendency to redirect attention towards the federation of specialised data stores and facilitating query processing across diverse data models [6]. This transition has the potential to yield numerous benefits: One key advantage of systems is their ability to effectively utilise different data models. This allows for enhanced semantic expressiveness in the underlying interfaces and the utilisation of the internal processing capabilities of component data stores. Furthermore, federated architectures provide query-specific data integration through the implementation of just-in-time transformation and migration processes. This approach holds promise for substantially diminishing operational complexity and overhead. Projects that are centred on the development of systems within this particular research domain arise from a multitude of different disciplines and tackle a wide range of topics. This diversity might pose challenges in terms of establishing a cohesive perspective on the body of work within this field [7].

The domain of data integration has had substantial growth over time, evolving from its initial purpose of offering a standardised query and update interface for structured databases within an organisation. It has now advanced to encompass the capability to search, exchange, and modify both structured and unstructured data, whether they are located within or outside the organisation. Two main challenges can arise while this field advances, One primary obstacle lies in the development of effective open-source tools to cater to various components inside data integration pipelines. One of the primary challenges lies in offering practitioners with feasible solutions to address the persistent issue of effectively integrating structured and unstructured data in a systematic manner [8].

Many different educational centers, universities and institutions are collecting data about their employees, students, and the education process. Each of these entities collect their data in different manners and in different areas of the process, however, their ultimate target is the same, to make the decision-making process easier. A system collecting these information and combining them together to provide better reporting capabilities and generating better knowledge about their for these different entities will make the process to go through much easier.

The primary objective of this proposed system is to address the challenge arising from the diverse and disparate information sources across various educational institutions. The research aims to develop a system that facilitates easy access to knowledge for all relevant stakeholders and emphasizes the significance of informed decision-making in these domains. This is particularly crucial in developing countries like Iraq, where universities lack efficient interconnectivity and are unable to leverage valuable information from one another to enhance the educational process based

DOI: http://dx.doi.org/10.25098/7.2.34

The Scientific Journal of Cihan University – Sulaimaniya PP: 271-291
Volume (7), Issue (2), December 2023
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

on the experiences of other institutions. The efficient utilization of useful information will be the aim and target for this research.

The proposed system is a web-based application that is making utilization of the 3-Tiered Architecture (3TA) (see Fig.1) which makes it efficient and easy to use while making sure all of its data is securely stored and kept at a hosted Database (DB)[9].

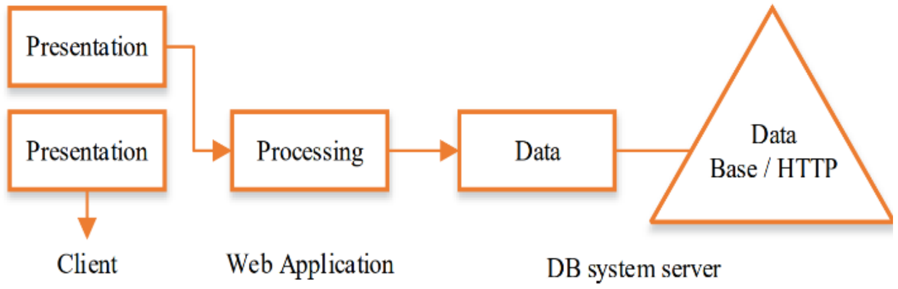


Fig. 1. Three-Tiered Architecture [9]

Additionally, Figure 2 below shows the overall functionality of our proposed system and how it is utilizing the 3TA:

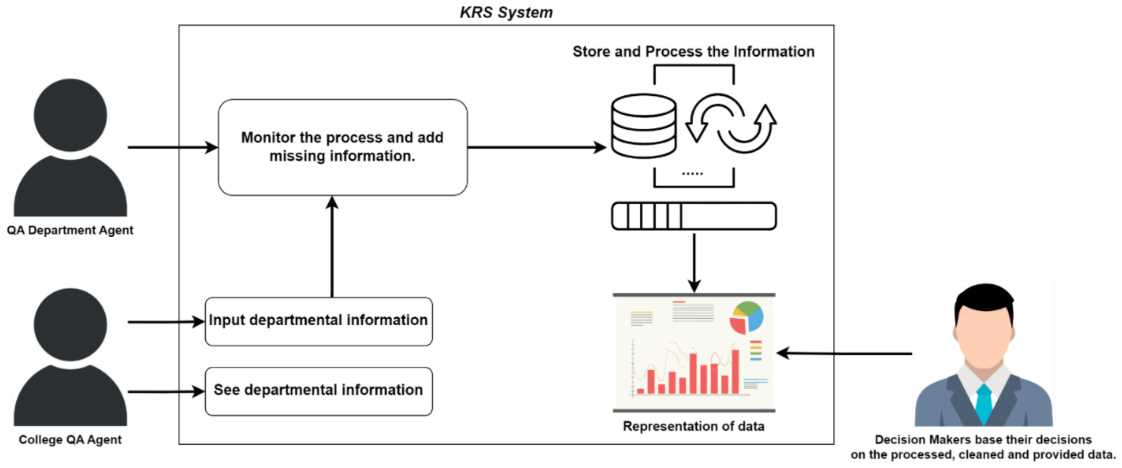


Fig. 2. The general structure of the proposed KRS system

As shown in Figure 2, the proposed system works exactly the same way that any 3TA application would work by gathering input from users, processing them, and then storing them inside databases, while also providing presentation of these data as well.

The paper's reminder is structured as follows: Related work is discussed in Section 2, in Section 3, the methods algorithms, and materials used are explained, and Section 4 includes the results of our work while also discussing them. Finally, in section 5, a conclusion of the subject is outlined.

DOI: http://dx.doi.org/10.25098/7.2.34

2. RELATED WORK

Knowledge representation is an essential part of a lot of fields in which all are using it in different ways or means to provide a better understanding of their fields and take necessary decisions to further advance their field, below a number of related works to knowledge representation and decision support systems are explored.

The study [1] by Solodovnikova et al. introduce a methodology for handling challenges related to evolving data sources and information requirements. They propose a data warehouse architecture that involves gathering data from various sources, using ETL procedures to standardize the data, and relying on metadata for system operation. The main focus of the study is on processing both detected and manually executed changes in data. The authors successfully applied this method to a publication data warehouse to demonstrate its effectiveness.

In the referenced study [10], researchers explore decision support systems and emphasize the importance of educational data mining, a growing research area focused on analyzing various types of data from educational settings. They introduced a new decision support tool designed to predict students' performance on year-end exams. This tool employs a hybrid prediction system that combines multiple machine learning algorithms, outperforming individual algorithms. Additionally, the solution is user-friendly and compatible with various platforms and operating systems. The research's goal is to improve student admission processes and enhance services in educational institutions.

Livieris et al. in their article [11], The paper provides an overview of Decision Support Systems and Learning Analytics and their applications in education. It outlines the characteristics and methods of these approaches. The paper introduces an integrated Decision Support System Model that primarily focuses on improving educational performance but can also be adapted for decision-making in other contexts. The model uses historical records, data analysis visualization systems, and integrates various data sources to present information in various visual formats to enhance decision-making efficiently.

In their publication [12], The authors, Myłka, Myłka, Kryza, and Kitowski, introduce the X2R system, which is designed to integrate data from various sources into an ontological knowledge base. The system's main goal is to provide a unified perspective on data stored in relational, XML, and LDAP sources within an organization. It achieves this by representing the data in RDF format, using a shared ontology, and ensuring compliance with predefined integrity requirements. The X2R platform enables users to specify complex data transformations between the original data sources and the unified knowledge base, allowing for the integration of diverse source schemas and target ontologies. The system also incorporates a range of integrity constraint primitives to maintain the quality and integrity of the unified dataset. Additionally, these constraints are utilized in a unique approach to enhance the semantic optimization of SPARQL queries.

In addition, Y. Li and H. Li [13] propose an architectural framework based on autoencoders for robust learning from streaming data. They utilize denoising autoencoders for online feature learning to improve the robustness of learned representations. To handle data shift between source and destination streaming data, they introduce an ensemble weighted technique effective in managing

DOI: http://dx.doi.org/10.25098/7.2.34

concept drift. They also create a transfer mechanism to convey label information across diverse domains. This integrated approach enhances the architecture's effectiveness in multi-stream classification problems, resulting in accurate predictions with limited labeled examples. Experiments on various datasets show that their method outperforms other relevant methodologies and achieves state-of-the-art results.

The paper [14], asserts Education plays a pivotal role in global development, making educational decision support crucial for students, educators, and institutions. The study proposes an educational decision support system designed for semester credit systems, incorporating educational data mining techniques to enhance its functionality. Challenges related to extracting information from this system are addressed, enabling administrators to gain practical insights from educational data. Knowledge-based decision support systems empower administrators to make informed decisions about students' academic progress and provide necessary support for successful graduation, helping to prevent wastage of effort, time, and financial resources for both students and educators.

The study[15], This scholarly endeavor empirically compares the use of different data sources, various classifiers, and ensemble methods for predicting student academic achievement. The study assesses the performance and efficiency of ensemble approaches using multiple data sources, as opposed to single data source base classifiers. The results demonstrate that combining diverse data sources and employing heterogeneous ensemble methods is highly effective and accurate in predicting student performance, as well as identifying students at risk of attrition.

In their study, WuHanrui, WuQingyao, and K NgMichael [16] propose a domain adaptation approach to improve learning in a target domain by utilizing information from a source domain. Transferring knowledge between these domains becomes more challenging when the data in both domains have different types of features, known as heterogeneous domain adaptation (HDA). To tackle this issue, they introduced a new framework called Knowledge Preserving and Distribution Alignment (KPDA). KPDA aims to create an extended target space while minimizing knowledge loss and maximizing domain distribution alignment. This is achieved by identifying a latent space using Laplacian graph terms and reconstruction regularization. Maximum Mean Discrepancy is used to align the distributions of source and target domains in the latent space. The mathematical formulation of KPDA involves a minimization problem with orthogonal constraints and two projection variables. The method employs the Gauss-Seidel iteration scheme, breaking the problem into two subproblems. These subproblems are then addressed using search algorithms that rely on the Barzilai-Borwein (BB) step size. The preliminary findings indicate the efficacy of the proposed approach.

B. Xu, S. Yan, S. Li, and Y. Du [17] in this study introduces an innovative architecture that employs federated transfer learning to classify students' grades while preserving privacy. It uses feature extractors and domain classifiers to align the feature distribution of participants in collaborative learning. The system ensures data confidentiality during transmission through homomorphic encryption and random masking. Experimental results show that knowledge transfer is efficient, even when dealing with partners having different data feature spaces and scales. Including target data parties with missing labels or limited features can enhance the training of student grade

DOI: http://dx.doi.org/10.25098/7.2.34

The Scientific Journal of Cihan University – Sulaimaniya **PP: 271-291**
Volume (7), Issue (2), December 2023
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

classification models, improving generalization and classification effectiveness for primary data participants.

In the study [18], H. Wu and M. K. Ng address the complexities of multi-source domain adaptation, particularly the challenges associated with diverse feature representations in distinct domains, known as Multi-source Heterogeneous Domain Adaptation (MHDA). They propose a novel approach called Multiple Graphs and Low-rank Embedding (MGLE) to capture the local structural characteristics of different domains using multiple graphs and learn low-rank embeddings of the target domain. MGLE further improves the acquired embedding by incorporating the original target data and includes domain discrepancy and domain relevance modules. They develop an iterative optimization approach to tackle these challenges. The study evaluates the effectiveness of MGLE on real-world datasets, and the results demonstrate that MGLE outperforms baseline approaches across various measures, including AUC, MAE, accuracy, precision, F1 score, and MCC, confirming its efficacy.

Although all the works explored in this section are somewhat close to our work, none of them has combined all the functionalities to design a system that utilizes academic knowledge to produce a decision-making support system for academic institutions to display their performance and also to compare it with other universities. The methodology of choice, approach and focused fields of the related works are outlined in Table 1 below:

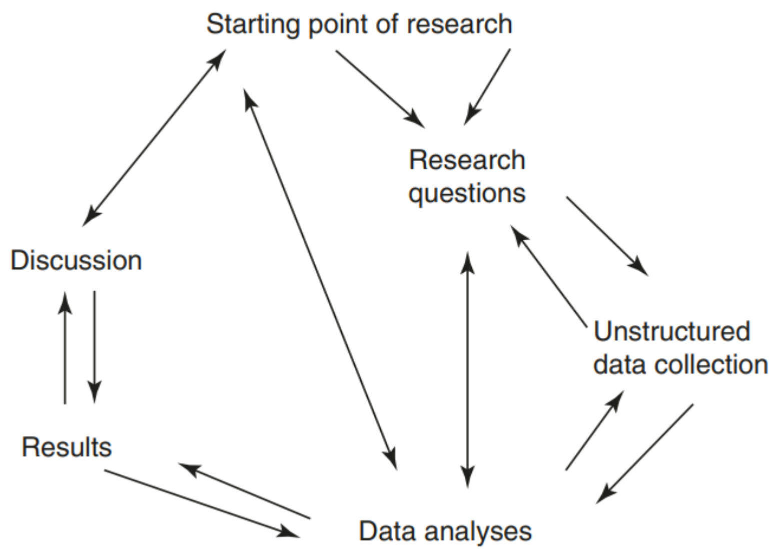


Fig. 3. The process of qualitative research [20]

The initial stage of qualitative research involves selecting a research topic, often initiated by a researcher's idea. At this point, there may be a lack of existing literature to support the concept being investigated. The research inquiries in qualitative research do not include hypotheses and are derived from the research's starting point. Data collection methods in qualitative research are open or semi-structured and can include interviews, observations, and written materials. Research questions can evolve during the study, which is common in exploratory research with open-ended data collection. Qualitative research standards suggest analyzing data as it is collected, allowing for adjustments to research questions. In this study, the research concept was deliberated upon, leading to the

DOI: http://dx.doi.org/10.25098/7.2.34

formulation of research inquiries. Data was collected, analyzed, and refined, with an iterative process. [20].

3.1 Fundamental Concept

In the discussion phase of qualitative research, the researcher goes beyond the raw data to interpret the findings and understand their significance. This involves a thorough analysis of the results in relation to the research questions or hypotheses. The interpretation is not a mere repetition of data but a detailed exploration of what the results mean in the research context.

Furthermore, in the discussion phase, the researcher compares and contrasts the findings with existing studies and theories, placing the research within a broader academic context. Practical, theoretical, or societal implications of the findings are examined, and recommendations for future research, policy changes, or professional practices may be suggested. It's important to acknowledge any study limitations, such as sample size, methodology, or data analysis issues. The discussion phase concludes with a summary of the findings and their implications, highlighting the research's contribution to the field.

The research questions phase in qualitative studies is a crucial step in defining the exploration or understanding goals of the research. Qualitative research questions are typically open-ended, broad, and aimed at understanding 'how' or 'why' something occurs, focusing on experiences, perspectives, or processes rather than hypothesis testing. Crafting effective research questions involves reviewing existing literature, ensuring they are original and significant, clear, focused, complex enough to merit detailed answers, and feasible within the study's scope.

These questions serve as a guide for the entire research process, influencing methodology, data collection, and analysis. They may evolve and be refined as the study progresses and more is learned about the topic.

3.2 Heterogeneous data acquisition

Unstructured data collection in qualitative research involves gathering information without a predefined format, offering flexibility to capture nuanced details, emotions, and complexities. Methods for unstructured data collection include in-depth interviews, which are open-ended conversations, observations, where researchers immerse themselves in the study setting, open-ended surveys and questionnaires that allow for participant responses in their own words, and content analysis, which analyzes unstructured data like text, images, or videos to identify themes and patterns. While unstructured data collection can be more time-consuming to collect and analyze than structured data, it provides a deeper and richer understanding of the research topic.

3.3 Data Test

In the data analysis phase of qualitative research, researchers interpret raw data to derive meaningful insights, aiming to answer the research question. This involves several steps, starting with data preparation, where collected data is organized and transcribed. Next comes coding, where labels (codes) are assigned to data segments representing various themes or patterns. The process identifies themes, which are then related to the research questions. Researchers interpret these themes, seeking

DOI: http://dx.doi.org/10.25098/7.2.34

to explain the 'how' and 'why' of phenomena. Validation methods are employed to check findings' reliability and validity, such as triangulation, member checking, or peer debriefing. The final step involves reporting findings in an organized manner, often using quotes from the data to illustrate the themes. The aim of qualitative data analysis is not to quantify or predict outcomes but to gain a deeper understanding of complex phenomena and explore diverse perspectives.

3.4 Software Implementation

The V-model software development methodology [21] was chosen for its suitability and is commonly used, especially in industries like automotive [22]–[24], due to its simplicity and well-defined documentation phases that aid in issue tracking during development. The V-model divides the system development process into two main categories: system definition and verification steps. In the system definition phase, developers establish system requirements based on high-level business needs, followed by high- and low-level design by system engineers. The implementation phase precedes the verification steps, which involve testing at both unit and system levels. These verification steps ensure that the implemented system's functionality aligns with the earlier-defined requirements. If any requirements are not met, adjustments to the design are made to meet all requirements. After system acceptance and release, ongoing operations and maintenance are conducted throughout the system's lifecycle.

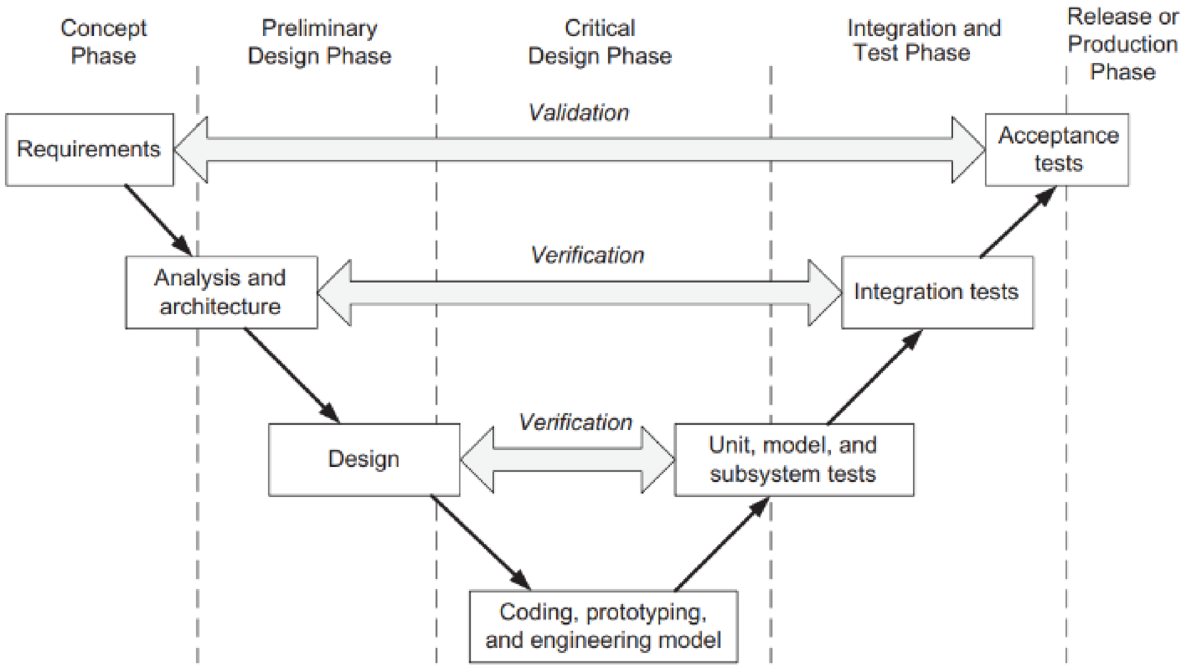


Fig 4. The V-model development process [25]

DOI: http://dx.doi.org/10.25098/7.2.34

3.5 Decision Making Algorithm

The Binary Decision Tree technique, specifically the J48 method, was chosen for decision-making in the proposed system because it can classify input data into discrete categories using a tree structure. This algorithm was selected for its simplicity and speed in the decision-making process. The J48 method employs a typical tree structure, including a root node, internal nodes, leaves, and branches, with each branch representing a sequence of nodes from the root to a leaf. Each node in this structure represents a specific trait or feature for classification. [26]. According to Mu et al. [27], Decision trees are widely used in data mining, spanning areas such as pattern recognition, machine learning, image processing, and information retrieval. They aim to uncover the underlying structure within data. In the context of particle classification, using a decision tree as the foundation holds promise due to its user-friendly and transparent nature. This approach can be advantageous for establishing thresholds based on morphological characteristics for particle classification.

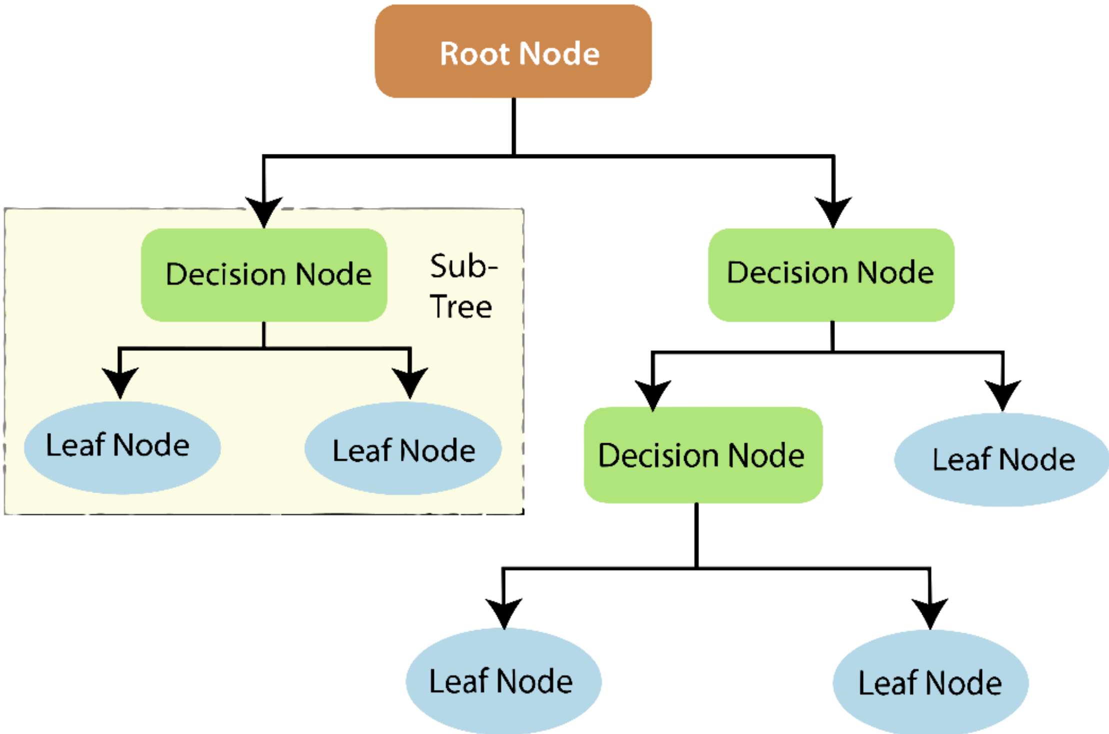


Fig 5. Decision Tree [28]

Figure 5 depicts the path that decision trees take so that it makes a specific decision. In our case, the details about each academic activity are categorized per academic year and stored within the database of the system. Later, when these data are needed to be used in a comparison to show the rating for each category for the university in case compared against other universities or compared with other academic years a simple (if and else) is used that translates the decision tree into programming language.

DOI: http://dx.doi.org/10.25098/7.2.34

The Scientific Journal of Cihan University – Sulaimaniya **PP: 271-291**
Volume (7), Issue (2), December 2023
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

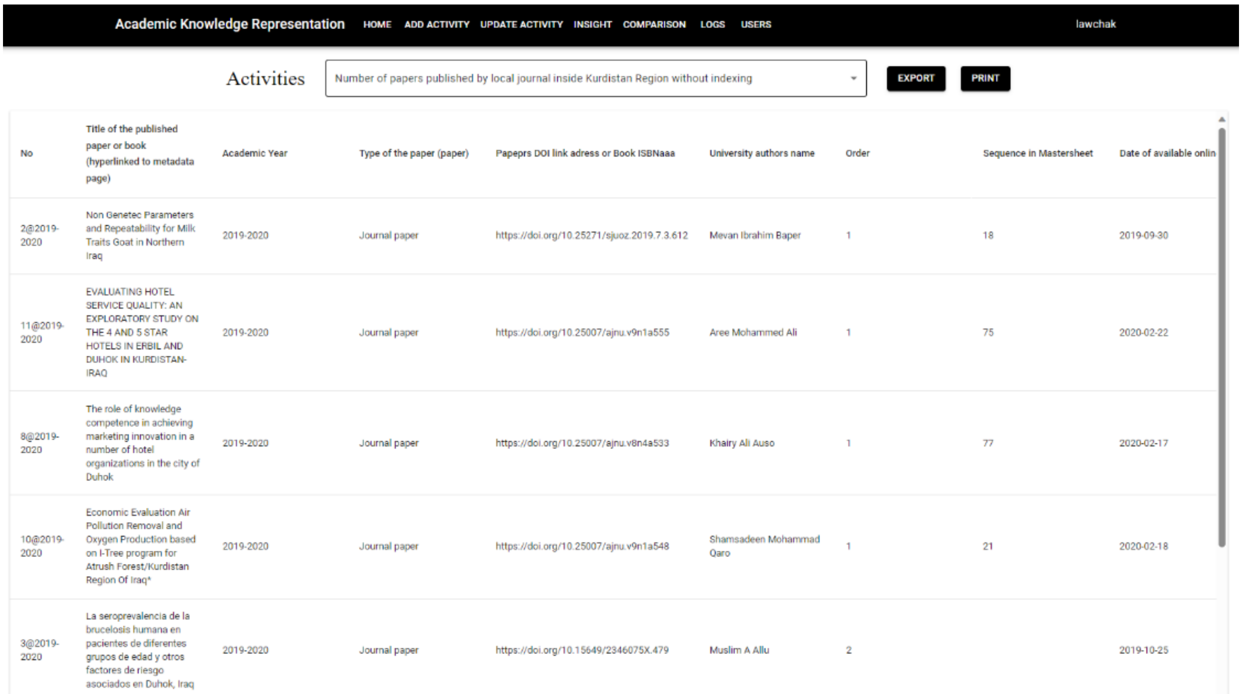
4. Evaluation and Results

Keeping the previously collected points in mind the proposed system has been developed and equipped with tools and methods to make it easy to use and detailed enough to help decision makers make the right decision and understand the processed information easily. In this section, the outcomes of the proposed system are discussed. Then, it shows the main interface of the aforementioned system and its various settings and feature set that allows a smooth operation of the Knowledge Representation process.

The main sections of the system are:

4.1 Home

The Home page is the main page of the system where user logs into, it provides a view to the stored information within the system regardless of the academic year.



No	Title of the published paper or book (hyperlinked to metadata page)	Academic Year	Type of the paper (paper)	Papeprs DOI link adress or Book ISBNaaa	University authors name	Order	Sequence in Mastersheet	Date of available onlin
2@2019-2020	Non Genetec Parameters and Repeatability for Milk Traits Goat in Northern Iraq	2019-2020	Journal paper	https://doi.org/10.25271/sjuoz.2019.7.3.612	Mevan Ibrahim Baper	1	18	2019-09-30
11@2019-2020	EVALUATING HOTEL SERVICE QUALITY: AN EXPLORATORY STUDY ON THE 4 AND 5 STAR HOTELS IN ERBIL AND DUHOK IN KURDISTAN-IRAQ	2019-2020	Journal paper	https://doi.org/10.25007/ajnu.v9n1a555	Aree Mohammed Ali	1	75	2020-02-22
8@2019-2020	The role of knowledge competence in achieving marketing innovation in a number of hotel organizations in the city of Duhok	2019-2020	Journal paper	https://doi.org/10.25007/ajnu.v8n4a533	Khairy Ali Auso	1	77	2020-02-17
10@2019-2020	Economic Evaluation Air Pollution Removal and Oxygen Production based on I-Tree program for Atrush Forest/Kurdistan Region Of Iraq*	2019-2020	Journal paper	https://doi.org/10.25007/ajnu.v9n1a548	Shamsadeen Mohammad Qaro	1	21	2020-02-18
3@2019-2020	La seroprevalencia de la brucelosis humana en pacientes de diferentes grupos de edad y otros factores de riesgo asociados en Duhok, Iraq	2019-2020	Journal paper	https://doi.org/10.15649/2346075X.479	Muslim A Allu	2		2019-10-25

Fig. 6. Home section in the AKR system

In addition to viewing the activities that are stored within the system as shown in Figure 6, user can also Print these information directly from the system and also Export NUR-Ready xlsx files that would make it easier for QA agents to insert these information directly into NUR google sheets.

4.2 Add Activities

In this section, the user is allowed to add new activities to each of the categories that is considered critical by the NUR ranking system. Users are allowed to add new activities either by typing down the details of them manually, or by adding them in bulk from Excel files (e.g., excel files from earlier years of NUR ranking system).

DOI: http://dx.doi.org/10.25098/7.2.34

The Scientific Journal of Cihan University – Sulaimaniya PP: 271-291
Volume (7), Issue (2), December 2023
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

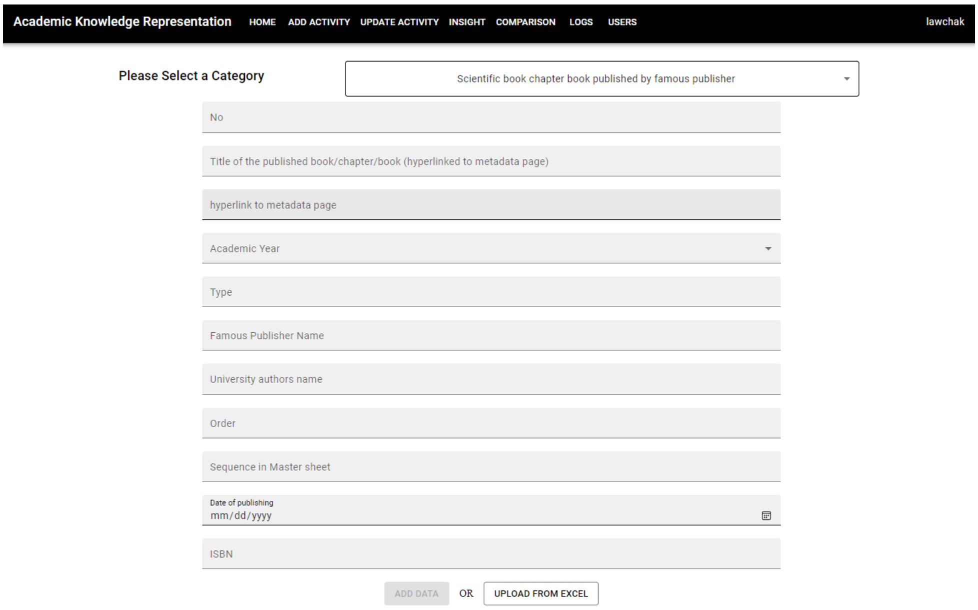


Fig. 7 Add Activity Section

As shown in Figure 7, the Add functionality allows users to type down the fields required for each category, while also allowing “Upload from excel” functionality.

4.3 Update Activities

The Update section allows the user to update the activities that are stored within the system.

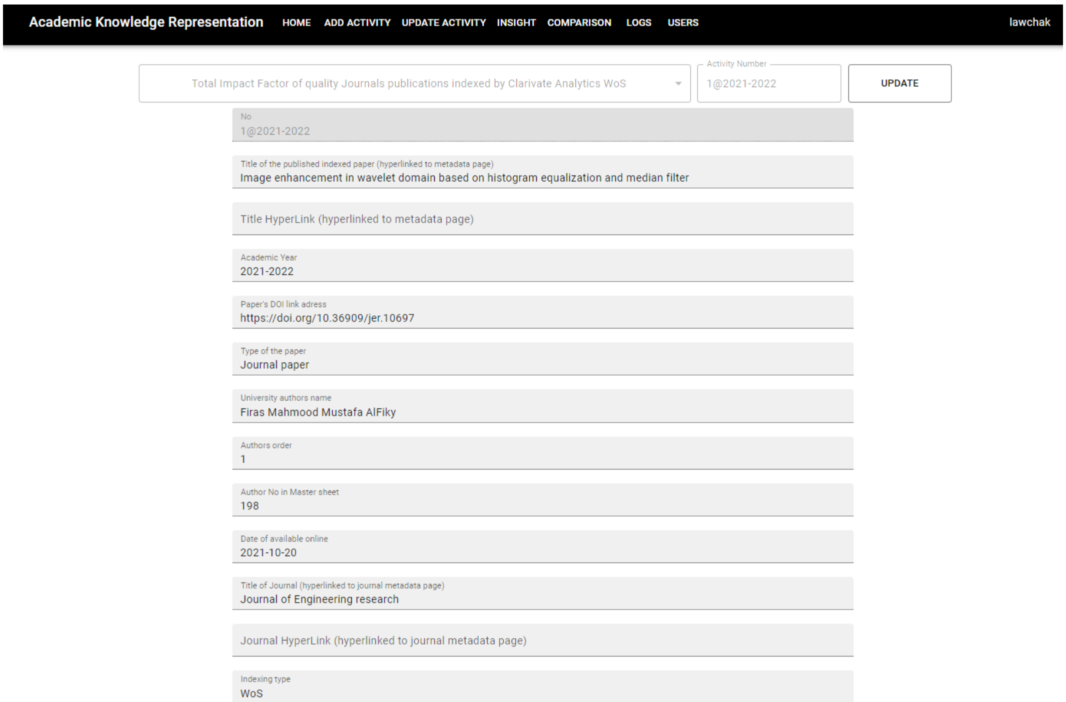


Fig. 8. Update Section in the AKR system

DOI: http://dx.doi.org/10.25098/7.2.34

The Scientific Journal of Cihan University – Sulaimaniya **PP: 271-291**
Volume (7), Issue (2), December 2023
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

As shown in Figure 8, the Update Activity section is quite simple which allows the user to select the required category, insert the Number of the item that needs to be updated and then simply modify any of the values as needed.

4.4 Insight

One of the most important sections of the proposed system is the Insight section. This section allows the user to show overall data and charts of the system and the information stored within the system.

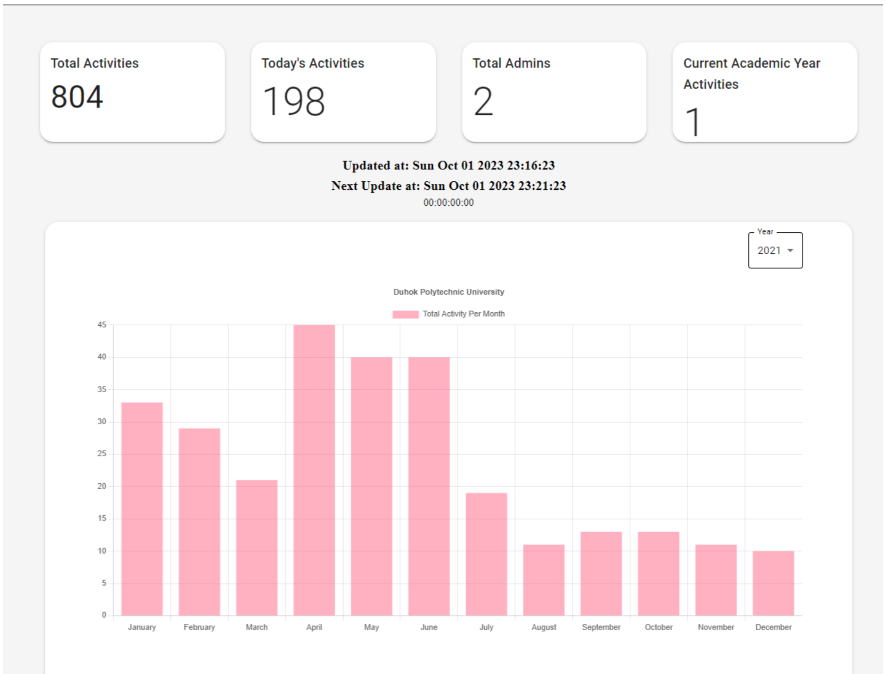


Fig. 9. Insight Section; Academic Activity per month graph

The Insight Section provides a chart as shown in Figure 9 that shows the number of activities per month during a specific academic year.

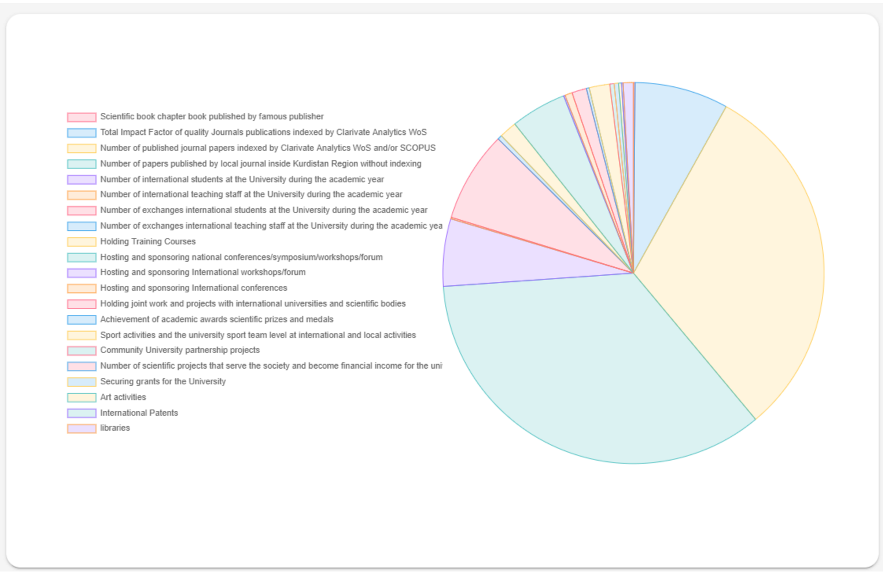


Fig. 10. Insight Section; Pie chart of types of activities

DOI: http://dx.doi.org/10.25098/7.2.34

Volume (7), Issue (2), December 2023
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

It also allows the user to see the data in a pie chart fashion which allows the user to see that distribution of activities per category. This provides a clearer understanding of the dominating activities within the university.

The section also provides a total count of activities per category (shown in Figure 11below) that is displayed in numbers, which is yet another feature to help understand the data better.

Number of each activity

libraries	7
Art activities	2
International Patents	1
Holding Training Courses	12
Securing grants for the University	3
Community University partnership projects	3
Hosting and sponsoring International conferences	5
Hosting and sponsoring International workshops/forum	1
Scientific book chapter book published by famous publisher	1
Achievement of academic awards scientific prizes and medals	2
Hosting and sponsoring national conferences/symposium/workshops/forum	38
Holding joint work and projects with international universities and scientific bodies	10
Sport activities and the university sport team level at international and local activities	14
Total Impact Factor of quality Journals publications indexed by Clarivate Analytics WoS	64
Number of papers published by local journal inside Kurdistan Region without indexing	281
Number of published journal papers indexed by Clarivate Analytics WoS and/or SCOPUS	248
Number of international students at the University during the academic year	46
Number of international teaching staff at the University during the academic year	1
Number of exchanges international students at the University during the academic year	62
Number of exchanges international teaching staff at the University during the academic year	3
Number of scientific projects that serve the society and become financial income for the university	0

Fig. 11. Insight Section; Total number of each activity

Finally, the Insight section also provides a Print function so that it could be easy to print out these information and make them readily available for reporting purposes.

4.5 Comparison

Throughout this section, the user is provided with a yet another important tool that is designed to compare the performance of a certain academic year of the university to other universities or to other (previous) academic years. This is also an addition to the importance of the proposed system as it provides a convenient comparison methodology for the university to be able to benchmark their own performance during the academic year or compare their performance to earlier years as shown in Figure 11.

DOI: http://dx.doi.org/10.25098/7.2.34

Volume (7), Issue (2), December 2023
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

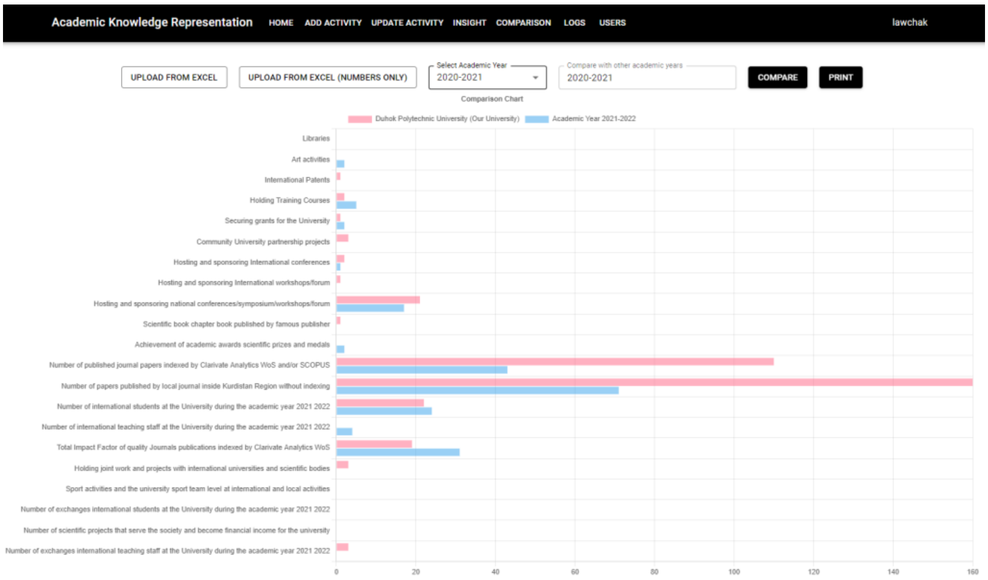


Fig. 12. Comparison Section; Comparing current academic year performance to earlier years.

As shown in Figure 12, the comparison shows the numbers of activities within a year and compares them directly to a different year, or to the numbers of a different university. If a university is willing to provide their NUR sheet for the year you can use that directly to compare to your university, otherwise, you can also compare only to the numbers of each activity. Also, the comparison section provides a "Rating" based on the number of each activity. The rating will show "Strong" if the number of a specific activity in our university is higher compared to the other university, "Weak" if it's the other way around, and "Moderate" if they are equal.

Rating

Activity	Our University	Compared University or Academic Year	Rating
Libraries	0	3	Weak
Art activities	0	9	Weak
International Patents	1	9	Weak
Holding Training Courses	2	8	Weak
Securing grants for the University	1	4	Weak
Community University partnership projects	3	7	Weak
Hosting and sponsoring International conferences	2	1	Strong
Hosting and sponsoring International workshops/forum	1	12	Weak
Hosting and sponsoring national conferences/symposium/workshops/forum	21	49	Weak
Scientific book chapter book published by famous publisher	1	0	Strong

Fig. 13. Comparison Section; Performance Rating

DOI: http://dx.doi.org/10.25098/7.2.34

Volume (7), Issue (2), December 2023
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

Figure 13 shows the Performance Rating in action when it is comparing the number of activities and rates them based on their numbers.

Finally, a descriptive text will be provided once a comparison is made in this section that is generated directly from ChatGPT AI Tool. This helps putting the numbers and charts into clear words that highlights the major differences in which each university has higher scores compared to the other.

AI Description

Based on the provided JSON file, let's analyze the data for the "Our University" and the "Compared Against University" in the 2nd and 3rd sections.

1. Libraries:
 - Our University: 0
 - Compared Against University: 2
 - Our University has significantly lower values than the Compared Against University.

2. Art activities:
 - Our University: 2
 - Compared Against University: 12
 - Our University has lower values than the Compared Against University.

3. International Patents:
 - Our University: 0
 - Compared Against University: 7
 - Our University has significantly lower values than the Compared Against University.

4. Holding Training Courses:
 - Our University: 5
 - Compared Against University: 5
 - Both universities have the same number of training courses.

5. Securing grants for the University:
 - Our University: 2
 - Compared Against University: 4

Fig. 14. shows the ChatGPT descriptive text in action when it is describing the facts about each provided category.

4.6 Logs

In this section, an overview of all the actions taken within the system is shown to the administrator of the system from adding activities, updating them, creating, or removing users and any other action that can be taken within the system.

4.7 Users

This section is also administrator-exclusive where it provides user management such as creating or removing users, updating their contact information or passwords and any other information related to the users of the system.

5. Validation: System Usability Score (SUS)

The System Usability Scale (SUS) is a standardized questionnaire that is widely used for assessing perceived usability. It was developed in the early 1980s as part of a usability engineering program. The SUS consists of 10 five-point items with alternating positive and negative tones. It is used to measure how users perceive the usability of computer systems they are working on. The SUS has gained popularity due to its simplicity and reliability. It has been widely adopted in industrial usability studies, accounting for 43% of post-study questionnaire usage. The paper that introduced the SUS has received over 5,600 citations on Google Scholar. Additionally, the SUS is freely available for

DOI: http://dx.doi.org/10.25098/7.2.34

use, with the only requirement being that users acknowledge the source of the measure in their reports and publications. Researchers and practitioners often choose to use the SUS because it provides a quick and effective way to measure perceived usability. Its widespread use and recognition as an "industry standard" contribute to its continued popularity in the field of usability assessment. [29]
The scoring process for the SUS involves several steps. Here is a breakdown of the process:

1. Converting raw item ratings into adjusted values: The initial stage involves the conversion of the unprocessed item scores into adjusted scores, which are commonly referred to as "score contributions." The adjusted scores span from 0, indicating the lowest rating, to 4, signifying the highest rating. The adjustment varies across items with odd and even numbers. To calculate the adjusted score for items with odd numbers, it is necessary to remove 1 from the raw value. To obtain the score for even-numbered items, decrease the raw score from 5.
2. Handling blank responses: If a respondent leaves an item blank, it should be given a raw score of 3, which represents the center of the five-point scale.
3. Compute the sum of adjusted scores: Add up the adjusted scores for the odd-numbered items and the adjusted scores for the even-numbered items separately.
4. Multiply by 2.5: Multiply the sum of the adjusted scores by 2.5 to obtain the standard SUS score.

To summarize, the formula for calculating the standard SUS score is as follows:

SUS = 2.5 * (20 + SUM(SUS01, SUS03, SUS05, SUS07, SUS09) - SUM(SUS02, SUS04, SUS06, SUS08, SUS10))

By following these steps and using the provided formula, you can calculate the final SUS score for a set of raw item ratings.

After the development of the system, the QA team was contacted to test the usability of the system and give an “SUS Score” as per their experience on using the system. This survey was taken internally within the QA related personnel within DPU university, and also externally within a number of universities within the KRI. The graph in Figure 15 shows the results of the survey taken from 15 participants in external universities that gave their score out of 1 to 5 to questions about the straightforwardness of the system, to the accuracy of the results displayed in the system. The average score of 80.5 out of 100 was achieved, shown in Figure 14 below.

Additionally, the same survey was conducted within the DPU that targeted the QA staff of multiple departments within the university and the QA department itself. This time the score was higher with the average score of 90, and this is due to the fact that these staff were the same ones interviewed before starting the development process. So, understanding and using the system was clearer. The result of this survey is shown in Figure 16.

DOI: http://dx.doi.org/10.25098/7.2.34

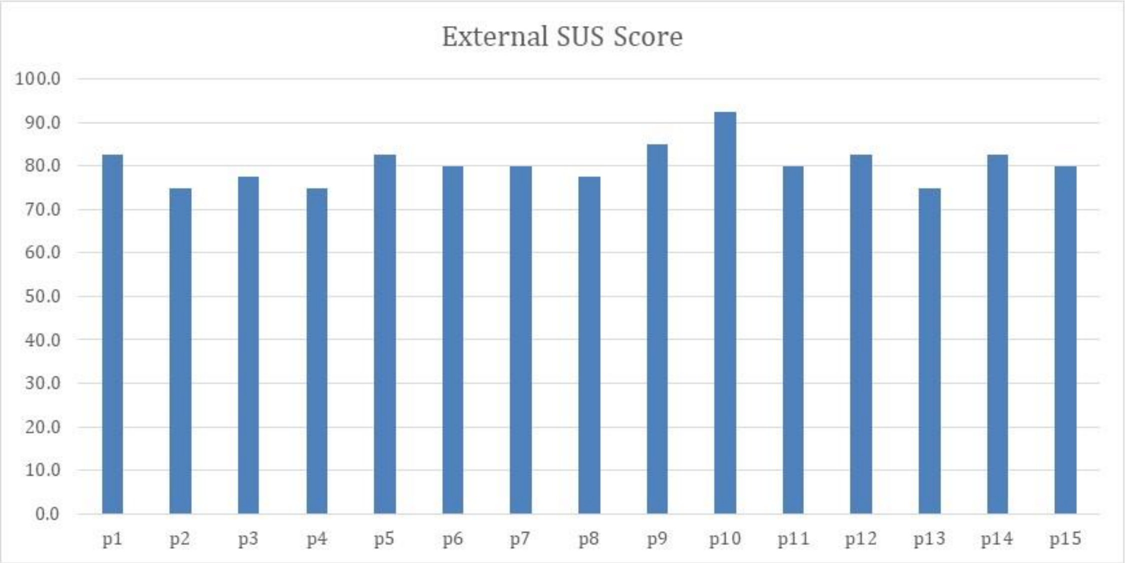


Fig. 15. External SUS score

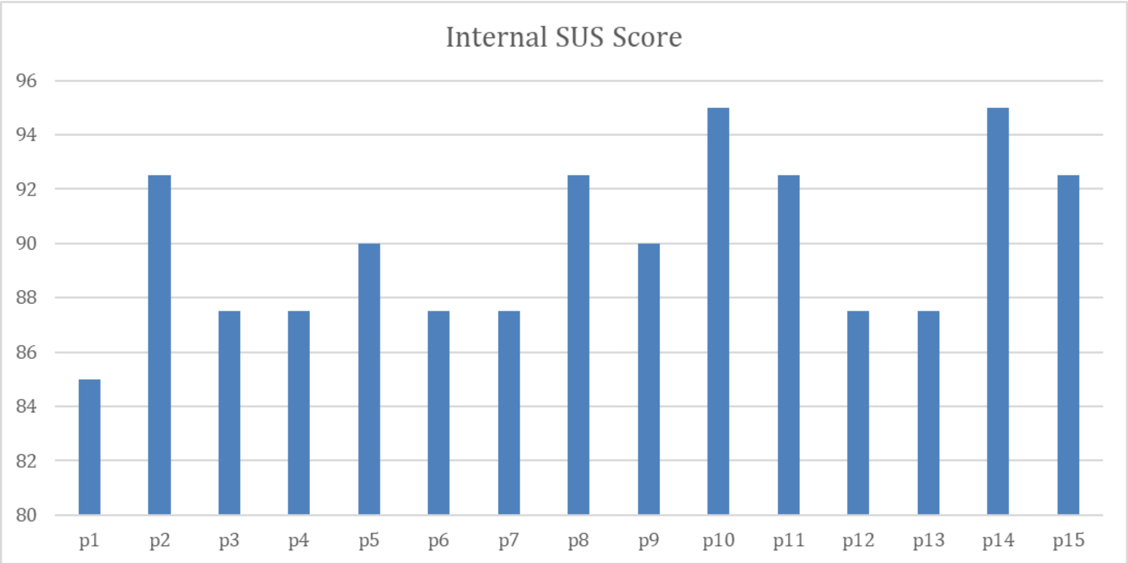


Fig. 16. Internal SUS score

DOI: http://dx.doi.org/10.25098/7.2.34

Conclusion

The academic field in the KRI is rapidly evolving; the types and numbers of academic activities are rising with it. These academic institutions all need a unified way to keep track of their progress and performance rating, hence, comes in the NUR system. However, the ways that are used to keep track of all the related academic activities performed by academic institutions are not unified, and mostly are not tracked even within a single academic institution. Our proposed system provides a unified way to store and keep track of all these data meanwhile providing real-time tracking and performance rating for institutions so that they could keep an eye on their performance in comparison to other institutions, or to their previous years. The AKR system also makes it easy for institutions that use it, to export and submit all their information regarding their academic activities to the NUR ranking system also. The SUS score that was gathered by the users of the system proves that the system is indeed working and makes the Knowledge representation process easier.

References

[1] D. Solodovnikova, L. Niedrite, and L. Svilpe, “Managing Evolution of Heterogeneous Data Sources of a Data Warehouse:,” in *Proceedings of the 23rd International Conference on Enterprise Information Systems*, Online Streaming, --- Select a Country ---: SCITEPRESS - Science and Technology Publications, 2021, pp. 105–117. doi: 10.5220/0010496601050117.

[2] Q. Zhang, D. Jian, R. Xu, W. Dai, and Y. Liu, “Integrating heterogeneous data sources for traffic flow prediction through extreme learning machine,” in *2017 IEEE International Conference on Big Data (Big Data)*, Boston, MA: IEEE, Dec. 2017, pp. 4189–4194. doi: 10.1109/BigData.2017.8258443.

[3] A. Carbonaro, F. Piccinini, and R. Reda, “Integrating Heterogeneous Data of Healthcare Devices to enable Domain Data Management,” *J. E-Learn. Knowl. Soc.*, vol. 14, no. 1, Jan. 2018, Accessed: Aug. 19, 2023. [Online]. Available: https://www.learntechlib.org/p/182316/

[4] K. Pal, “Integrating Heterogeneous Enterprise Data Using Ontology in Supply Chain Management,” in *Big Data and Knowledge Sharing in Virtual Organizations*, IGI Global, 2019, pp. 71–102. doi: 10.4018/978-1-5225-7519-1.ch003.

[5] A. E. W. Johnson *et al.*, “MIMIC-III, a freely accessible critical care database,” *Sci. Data*, vol. 3, no. 1, p. 160035, May 2016, doi: 10.1038/sdata.2016.35.

[6] C. R. on July 14 and 2015, “The Case for Polystores – ACM SIGMOD Blog.” Accessed: Aug. 19, 2023. [Online]. Available: http://wp.sigmod.org/?p=1629

[7] R. Tan, R. Chirkova, V. Gadepally, and T. G. Mattson, “Enabling query processing across heterogeneous data models: A survey,” in *2017 IEEE International Conference on Big Data (Big Data)*, Boston, MA: IEEE, Dec. 2017, pp. 3211–3220. doi: 10.1109/BigData.2017.8258302.

[8] B. Golshan, A. Halevy, G. Mihaila, and W.-C. Tan, “Data Integration: After the Teenage Years,” in *Proceedings of the 36th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems*, Chicago Illinois USA: ACM, May 2017, pp. 101–106. doi: 10.1145/3034786.3056124.

DOI: http://dx.doi.org/10.25098/7.2.34

[9] B. R. Ibrahim, S. R. M. Zeebaree, and B. Kh. Hussan, "Performance Measurement for Distributed Systems using 2TA and 3TA based on OPNET Principles," *Sci. J. Univ. Zakho*, vol. 7, no. 2, Jun. 2019, doi: 10.25271/sjuoz.2019.7.2.603.

[10] J.-F. Chen and Q. H. Do, "TRAINING NEURAL NETWORKS TO PREDICT STUDENT ACADEMIC PERFORMANCE: A COMPARISON OF CUCKOO SEARCH AND GRAVITATIONAL SEARCH ALGORITHMS," *Int. J. Comput. Intell. Appl.*, vol. 13, no. 01, p. 1450005, Mar. 2014, doi: 10.1142/S1469026814500059.

[11] I. Livieris, T. Mikropoulos, and P. Pintelas, "A decision support system for predicting students' performance," *Themes Sci. Technol. Educ.*, vol. 9, no. 1, pp. 43–57, Jul. 2016.

[12] "[PDF] Towards a Decision Support System based on Learning Analytics - Free Download PDF." Accessed: Aug. 19, 2023. [Online]. Available: https://silo.tips/download/towards-a-decision-support-system-based-on-learning-analytics

[13] A. Myłka, A. Myłka, B. Kryza, and J. Kitowski, "Integration of Heterogeneous Data Sources in an Ontological Knowledge Base," *Comput. Inform.*, vol. 31, no. 1, Art. no. 1, 2012.

[14] Y. Li and H. Li, "Online transferable representation with heterogeneous sources," *Appl. Intell.*, vol. 50, no. 6, pp. 1674–1686, Jun. 2020, doi: 10.1007/s10489-019-01620-3.

[15] T. N. C. Vo and H. P. Nguyen, "A Knowledge-Driven Educational Decision Support System," in *2012 IEEE RIVF International Conference on Computing & Communication Technologies, Research, Innovation, and Vision for the Future*, Ho Chi Minh City, Vietnam: IEEE, Feb. 2012, pp. 1–6. doi: 10.1109/rivf.2012.6169819.

[16] O. W. Adejo and T. Connolly, "Predicting student academic performance using multi-model heterogeneous ensemble approach," *J. Appl. Res. High. Educ.*, vol. 10, no. 1, pp. 61–75, Jan. 2018, doi: 10.1108/JARHE-09-2017-0113.

[17] H. Wu, Q. Wu, and M. K. Ng, "Knowledge Preserving and Distribution Alignment for Heterogeneous Domain Adaptation," *ACM Trans. Inf. Syst.*, vol. 40, no. 1, p. 16:1-16:29, Sep. 2021, doi: 10.1145/3469856.

[18] B. Xu, S. Yan, S. Li, and Y. Du, "A Federated Transfer Learning Framework Based on Heterogeneous Domain Adaptation for Students' Grades Classification," *Appl. Sci.*, vol. 12, no. 21, Art. no. 21, Jan. 2022, doi: 10.3390/app122110711.

[19] H. Nassaji, "Good qualitative research," *Lang. Teach. Res.*, vol. 24, no. 4, pp. 427–431, Jul. 2020, doi: 10.1177/1362168820941288.

[20] H. Kyngäs, "Qualitative Research and Content Analysis," in *The Application of Content Analysis in Nursing Science Research*, H. Kyngäs, K. Mikkonen, and M. Kääriäinen, Eds., Cham: Springer International Publishing, 2020, pp. 3–11. doi: 10.1007/978-3-030-30199-6_1.

[21] A. Al-Momani, F. Kargl, R. Schmidt, A. Kung, and C. Bosch, "A Privacy-Aware V-Model for Software Development," in *2019 IEEE Security and Privacy Workshops (SPW)*, San Francisco, CA, USA: IEEE, May 2019, pp. 100–104. doi: 10.1109/SPW.2019.00028.

[22] M. Kuhrmann, D. Niebuhr, and A. Rausch, "Application of the V-Modell XT – Report from a Pilot Project," in *Unifying the Software Process Spectrum*, M. Li, B. Boehm, and L. J. Osterweil, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2006, pp. 463–473. doi: 10.1007/11608035_37.

DOI: http://dx.doi.org/10.25098/7.2.34

[23] R. Rana, M. Staron, C. Berger, J. Hansson, M. Nilsson, and F. Törner, "Increasing Efficiency of ISO 26262 Verification and Validation by Combining Fault Injection and Mutation Testing with Model based Development," presented at the International Conference on Software Engineering and Applications, SCITEPRESS, Jul. 2013, pp. 251–257. doi: 10.5220/0004592002510257.

[24] M. Broy, "Challenges in automotive software engineering," in *Proceedings of the 28th international conference on Software engineering*, in ICSE '06. New York, NY, USA: Association for Computing Machinery, May 2006, pp. 33–42. doi: 10.1145/1134285.1134292.

[25] K. R. Fowler, "Introduction to Good Development," in *Developing and Managing Embedded Systems and Products*, Elsevier, 2015, pp. 1–38. doi: 10.1016/B978-0-12-405879-8.00001-5.

[26] C. K. Madhusudana, H. Kumar, and S. Narendranath, "Fault Diagnosis of Face Milling Tool using Decision Tree and Sound Signal," *Mater. Today Proc.*, vol. 5, no. 5, pp. 12035–12044, 2018, doi: 10.1016/j.matpr.2018.02.178.

[27] Y. Mu, X. Liu, and L. Wang, "A Pearson's correlation coefficient based decision tree and its parallel implementation," *Inf. Sci.*, vol. 435, pp. 40–58, Apr. 2018, doi: 10.1016/j.ins.2017.12.059.

[28] B. Charbuty and A. Abdulazeez, "Classification Based on Decision Tree Algorithm for Machine Learning," *J. Appl. Sci. Technol. Trends*, vol. 2, no. 01, pp. 20–28, Mar. 2021, doi: 10.38094/jastt20165.

[29] J. R. Lewis, "The System Usability Scale: Past, Present, and Future," *Int. J. Human–Computer Interact.*, vol. 34, no. 7, pp. 577–590, Jul. 2018, doi: 10.1080/10447318.2018.1455307.