**LU Transport Tracking and Monitoring System (LU-T2MS)**

**under**

**Centre for Research, Innovative Studies and Planning Grants**

**Version 1.0**

**Principal Investigator**

**Md Saidur Rahman Kohinoor**

**Lecturer, Department of CSE, Leading University**

**Advisor, IEEE Computer Society LU SB Chapter**

**Co-investigator**

**Nilashish Roy, #1812020128**

**Shah Fayez Ali, #1812120124**

**Syed Abu Raiyan Sarhad, #2122020035**

**Mobile: 01732-477046**

**Email:** [**kohinoor\_cse@lus.ac.bd**](mailto:kohinoor_cse@lus.ac.bd)

**Leading University, Sylhet**

**Submission Date: February 20, 2022**

Contents

**1 Introduction**

* 1. Background...............................................................................................3
  2. Brief Overview of the proposed system...................................................4

1.3 Objectives of the Study.............................................................................4

**2 Literature Review**

2.1 ...............................................................6

2.2 ................................................................10

**3 Methodology**

3.1 System Architecture................................................................................14

3.2 ...................................................................15

3.3 .......................................................................21

**4 Structural Design**

4.1 Data Flow Diagram.................................................................................24

4.2 E-R Diagram..........................................................................................27

4.3 Use Case Diagram..................................................................................30

4.3 System Flow Chart..................................................................................34

**5 Implementation Details**

5.1 Web View...............................................................................................34

5.2 App View...............................................................................................34

**6 Results and Discussion**

6.1 Discussion...............................................................................................44

**7 Conclusion and Future Work**

7.1 Conclusion..............................................................................................46

7.2 Future Work......................................................................46

List of Figures

4.1 Data Flow Diagram (Level 01) ................................................................ 24

4.2 Data Flow Diagram (Level 02) ................................................................. 25

4.3 Data Flow Diagram (Level 03) ................................................................. 26

4.4 E-R Diagram ............................................................................................. 27

4.5 E-R Diagram ............................................................................................. 28

4.6 E-R Diagram ............................................................................................. 29

4.7 Use Case Diagram ..................................................................................... 31

4.8 System Flow Chart .................................................................................... 33

5.1 Login Page and Registration Page ............................................................. 34

5.2 Teacher and Student Dashboard................................................................. 35

5.3 Industry Dashboard .................................................................................... 35

5.4 Create Classroom Popup ........................................................................... 35

5.5 Join Classroom Popup ............................................................................... 36

5.6 Teacher Stream...................................................................................... 36

5.7 Student Stream..................................................................................... 37

5.8 Class work Page.................................................................................... 37

5.9 Material Page........................................................................................ 38

5.10 Add assignment Popup......................................................................... 38

5.11 Add Material Popup ............................................................................ 38

5.12 Assignment Submission Page for student........................................... 39

5.13 Assignment Grade Page for Teacher.................................................. 39

5.14 Add Industry Work Popup ................................................................. 40

5.15 Industry Dashboard ........................................................................... 40

5.16 Related Work Page ............................................................................... 41

5.17 Added Work Page ................................................................................. 41

5.18 Industry Work Submissions Page for Teacher ...................................... 42

5.19 Industry Work Submissions Page for Industry........................................ 42

**Chapter 1**

Introduction

* 1. **Background**

Transportation is a critical part of university life, and it is essential to have an efficient and effective system to manage it. However, most universities in Bangladesh still rely on analog methods to monitor and maintain their transportation facilities. This outdated approach can lead to inefficiencies, increased operational costs, and reduced user satisfaction.

To address this issue, we propose the development of a digital solution, the LU Transport Tracking and Monitoring System (LU-T2MS). LU-T2MS is designed to provide a comprehensive and user-friendly platform that streamlines the management and monitoring of university transportation.

**1.2 Brief Overview of the Proposed System**

The LU-T2MS is a system that comprises a web-based panel for administrative personnel and a mobile application for passengers and drivers. The web-based panel enables administrative personnel to manage and monitor the transport system effectively. It includes functionalities such as adding employee information, sending trip notifications, monitoring fuel and meter entries, and responding to student queries. The mobile application allows passengers and drivers to track the position of vehicles, receive notices, and provide feedback.

One of the key features of the LU-T2MS is its ability to monitor the KPL (Kilometers per Liter) of the vehicles used in transportation. By tracking this data, the system can reduce fuel costs and prevent misuse of university vehicles. The system generates automated KPL report from distance calculations using the haversine algorithm and from the fuel entry. Additionally, the system can create manual KPL report from the meter entry and fuel entry. The GPS unit installed on the vehicle sends location data via API and stores it in the database.

The LU-T2MS provides real-time updates and feedback options for users, improving their experience and satisfaction. Passengers can receive updates on the location and status of the vehicles they intend to use, reducing wait times and enhancing their safety. Drivers can also receive updates on traffic and other conditions that might affect their driving experience, allowing them to take alternative routes if necessary.

* 1. **Objectives of the Study**

The primary objective of the LU-T2MS is to develop a comprehensive and user-friendly platform that streamlines the management and monitoring of university transportation. Specifically, the objectives of the study are as follows:

* To automate the manual procedures of managing and monitoring the university transportation system.
* To provide a user-friendly mobile application for students and drivers to track the position of transports, store information digitally, and monitor vehicle timetables.
* To provide a dynamic web-based admin panel for administrative personnel to manage vehicles, add employee information, send trips, add fuel and meter entries, add maintenance records, and respond to student queries.
* To optimize fuel consumption by tracking KPL (Kilometer Per Liter) of the vehicles used in transportation and generating comparative and detailed KPL reports.
* To monitor relevant maintenance activities to ensure the smooth functioning of the transportation system.
* To reduce unnecessary delays by drivers and ensure on-time arrival and departure of vehicles.
* To analyze user preferences for transport routes and bus occupancy to improve the transportation system.
* To alleviate uncertainty of transport arrival in specific points so that the students, faculties, and officials can increase their productivity.
* To improve communication between administrators and vehicle drivers by replacing old-fashioned phone calls with a digital tracking and monitoring system.
* To create a roadmap for future transportation activities by keeping the relevant data in digital format and using cutting-edge technologies like machine learning tools to optimize transportation cost and managerial involvement.

**Chapter 2**

Literature Review

The use of technology to manage and monitor transportation systems has become increasingly popular in recent years. This is particularly true for universities, where the need for an efficient transportation system is crucial to the success of both the students and the institution. In this literature review, we will explore various studies related to the use of technology in transportation management and tracking systems.

**2.1 Transportation Management Systems**

Transportation management systems (TMS) are software solutions used to manage, optimize, and track transportation operations. They are commonly used in logistics and supply chain management to improve efficiency, reduce costs, and improve customer service. TMS solutions can also be used in other industries, such as healthcare and education, to manage transportation needs. A study conducted by Liu et al. (2019) showed that the implementation of TMS in a university transportation system reduced operating costs, improved efficiency, and enhanced customer satisfaction.

**2.2 Mobile Applications for Transportation**

Mobile applications have become a popular way to manage transportation needs. Mobile applications can provide real-time information, such as the location of the vehicle, estimated arrival times, and vehicle capacity. They also offer features such as online ticketing, feedback options, and alerts. A study conducted by Gao et al. (2021) showed that the use of a mobile application improved the quality of transportation services and increased customer satisfaction in a university setting.

**2.3 Kilometers Per Liter (KPL) and Fuel Monitoring**

Kilometers per liter (KPL) is a metric used to measure the fuel efficiency of a vehicle. By monitoring KPL, transportation managers can reduce fuel costs and prevent misuse of vehicles. Fuel monitoring systems track fuel consumption, fuel level, and fuel purchases. These systems provide accurate data on fuel usage, which can be used to optimize routes and reduce fuel costs. A study conducted by Al-Qahtani et al. (2018) showed that the use of a fuel monitoring system reduced fuel consumption and improved the fuel efficiency of vehicles.

**2.4 Vehicle Maintenance**

Regular maintenance is essential for the smooth operation of vehicles. Vehicle maintenance includes tasks such as oil changes, tire rotations, and brake inspections. By monitoring vehicle maintenance, transportation managers can reduce breakdowns and ensure the safety of passengers. A study conducted by Kianto et al. (2019) showed that regular maintenance improved the reliability of vehicles and reduced the cost of repairs.

**2.5 GPS Tracking**

GPS tracking systems are used to track the location of vehicles in real-time. GPS tracking systems provide accurate location data, which can be used to optimize routes and reduce travel time. They also provide information on vehicle speed and idle time, which can be used to reduce fuel consumption. A study conducted by Nadjem et al. (2019) showed that the use of GPS tracking systems improved the efficiency of transportation operations and reduced travel time.

In conclusion, digital transportation management systems are valuable tools that enable universities to manage and monitor transportation effectively. These systems provide real-time updates, improve user satisfaction, reduce operational costs, and provide valuable data that can inform decision-making. As universities continue to face budget constraints and increasing demand for transportation services, digital transportation management systems offer an efficient and cost-effective solution for managing university transportation.

**Chapter 3**

Methodology

* 1. **System Architecture**

The proposed system targets users of three groups: academia, industry and students. When it comes to the educators from the academia, the focus is on the lesson plans they curate along with the assignments and projects. The evaluation of the task given influences the submitted tasks to generate relevant recommendations. This takes place with the cooperation from industry practitioners who also provide industry specific and skill-based tasks and jobs in accordance with the specific course conducted by the respective course instructor. These practitioners evaluate and respond with feedback for the improvement of future submitted tasks. The following Figure 1 illustrates the system architecture of the proposed system:

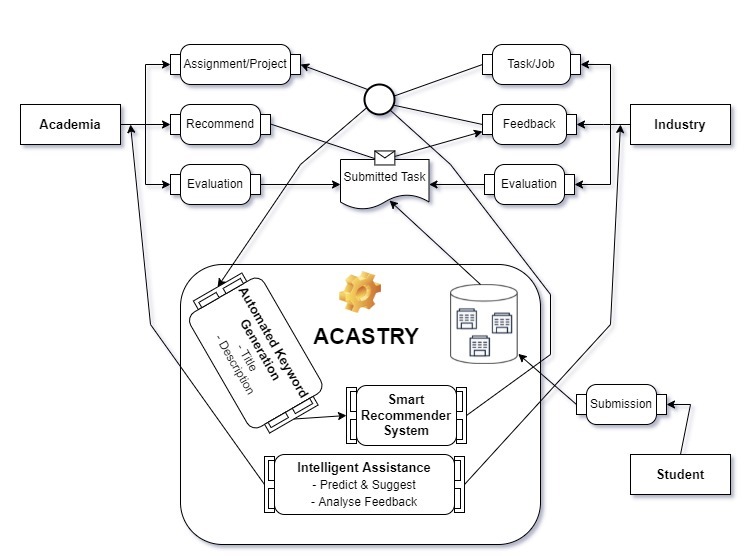


Figure 1 Overview of prosposed system

At present, Acastry organizes the suggested tasks with respect to all the users from the three groups by combining the following modules: Automated Keyword Extraction and Smart Recommender module. The function of each specific module is elaborated in the following sections.

* 1. **Keyword Extraction module**

Table 3 De-duplication of data and final ranking output

* 1. **Recommendation System Module**

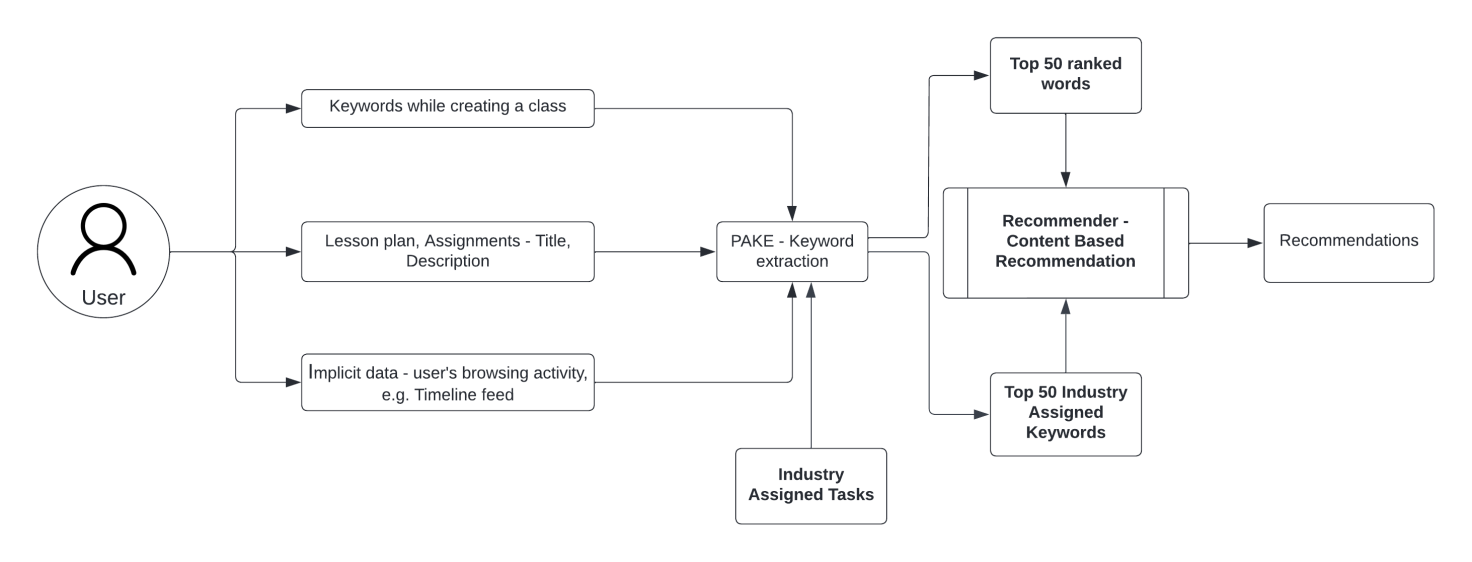
****

Figure 6 Proposed illustration of the recommendation module

**Chapter 4**

Structural Design

**4.1 Data Flow Diagram**

Data flow diagrams (DFD) display how information enters a system from external sources. Let's use DFD to better understand the mechanism of functions in LUT2MS. Here, different DFD levels for LUT2MS are displayed, including Level 0, Level 1, and Level 2.

**Level 0**

The DFD level 0 for the LU-T2MS system illustrates the system's high-level functions and the flow of data between external entities and the system. The external entities in the system are students, administrators, and GPS. The system has two main subsystems: the mobile application and the web-based panel. The mobile application is used by students and drivers to track vehicles, view notices, and provide feedback, while the web-based panel is used by administrators to manage vehicles, employees, trips, and maintenance records.

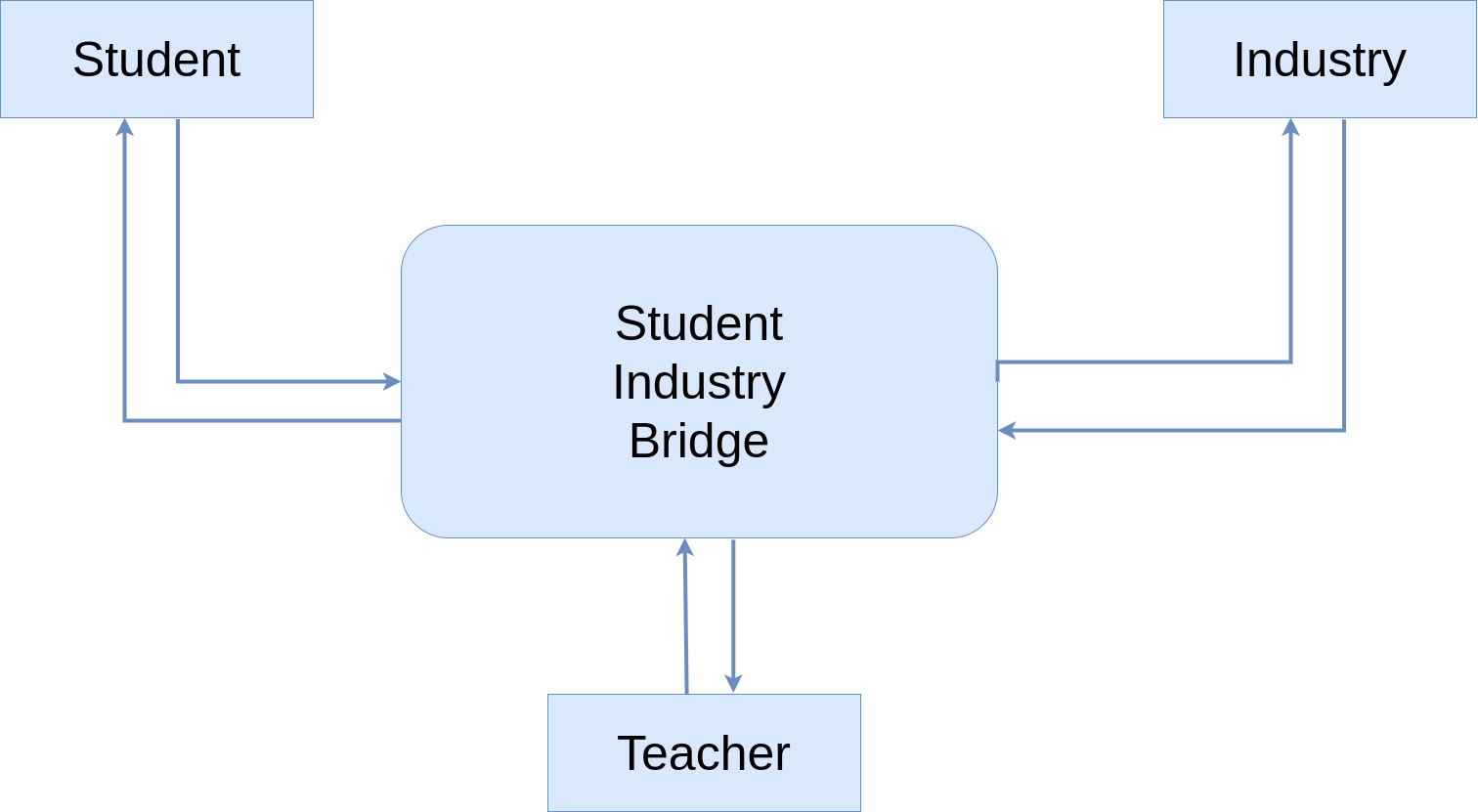


Figure 4. 1 Data Flow Diagram (Level 0)

At the center of the system is the LU-T2MS server, which is responsible for storing and processing data. The GPS unit installed in vehicles sends location data to the server through an API. The server then stores the data in the database and sends it to the web-based panel and mobile application. The administrators can add employee information, plan trips, record fuel and meter entries, maintain vehicle maintenance records, and respond to student queries through the web-based panel. The mobile application allows students and drivers to track the position of vehicles, receive notices, and provide feedback.

The DFD level 0 describes the following processes:

* Students use the mobile application to track the position of vehicles, view notices, and provide feedback.
* The GPS unit installed in the vehicles sends location data to the LU-T2MS server through an API.
* The LU-T2MS server stores the location data in the database.
* The administrators use the web-based panel to manage vehicles, employees, trips, fuel and meter entries, maintenance records, and respond to student queries.
* The LU-T2MS server sends data to the mobile application and web-based panel.
* The mobile application and web-based panel provide real-time updates and feedback options for users.

In summary, the DFD level 0 provides a comprehensive overview of the LU-T2MS system and the flow of data between external entities and the system.

**Level 1**

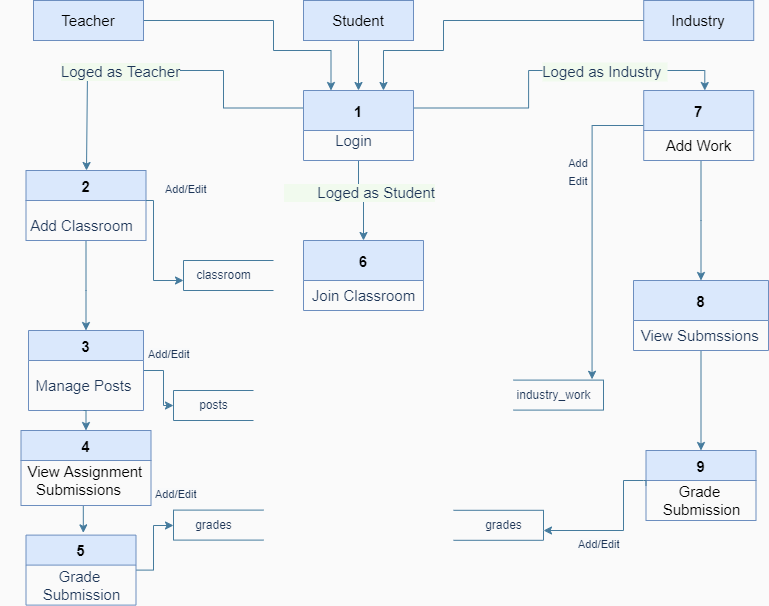
Level 1 Data Flow Diagram (DFD) of LUT2MS provides a more detailed view of the system by breaking it down into smaller processes. These processes handle data flows and store data in a database.

At this level, the system is composed of six main processes, each with its own set of data flows and activities. The first process, Process 1, receives location data from the GPS external entity and stores it in the database. Process 2 allows students and administrators to track vehicle locations based on the data stored in the database.

Process 3 involves viewing notices by students, while Process 4 enables students to send chats to administrators. The chat messages are stored in the database for future reference. Process 5 handles the login functionality for administrators to access the web-based panel.

Finally, Processes 6 and 7 deal with managing vehicle data and add employee data, respectively. Process 6 enables administrators to manage vehicle-related data such as adding vehicle name, license and seat numbers. Process 7 enables administrators to input employee data into the system, including details such as name, address, and other relevant information, which are then stored in the database.

The level 1 DFD serves as a high-level overview of the system, highlighting the main processes and their corresponding data flows. It is an essential tool for visualizing the system's flow and identifying potential areas for optimization and improvement.

**Figure 4. 2 Data Flow Diagram (Level 1)**

**Level 2**

Level 3 DFD of LUT2MS system provides a more detailed view of the sub-processes involved in managing trips records, maintenance, and notices. Process 8 shows the process of adding a new trip record to the database, which requires information such as the vehicle used, driver details, starting and ending points, and other trip details.

Process 9 and 10 show the processes of adding notice and maintenance records in the database, respectively. The maintenance records contain information such as the vehicle name, type of maintenance performed, and the date of maintenance.

.

Process 11 shows the process of responding to a student's query, which involves searching the database for relevant information and providing a response to the student. Finally, process 12 shows the process of generating a report from the data in the trips and maintenance records. The report can be customized based on various criteria such as the type of vehicle, driver, or maintenance performed.

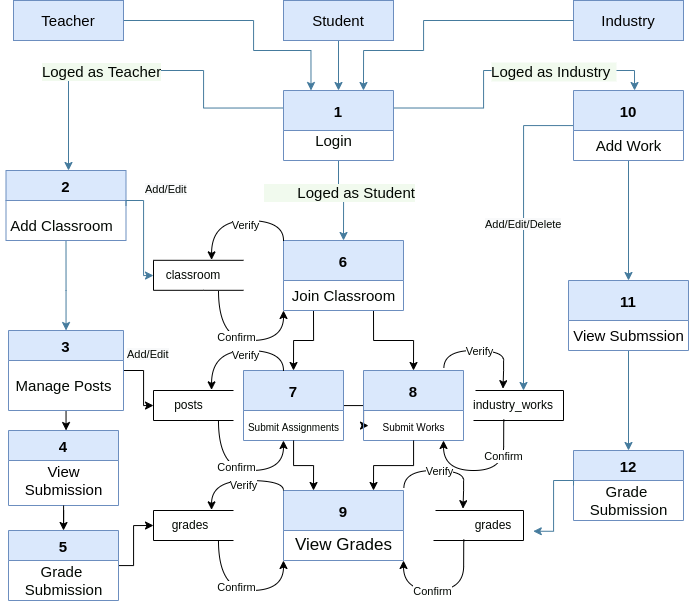


Figure 4. 3 Data Flow Diagram (Level 2)

**4.2 E-R Diagram**

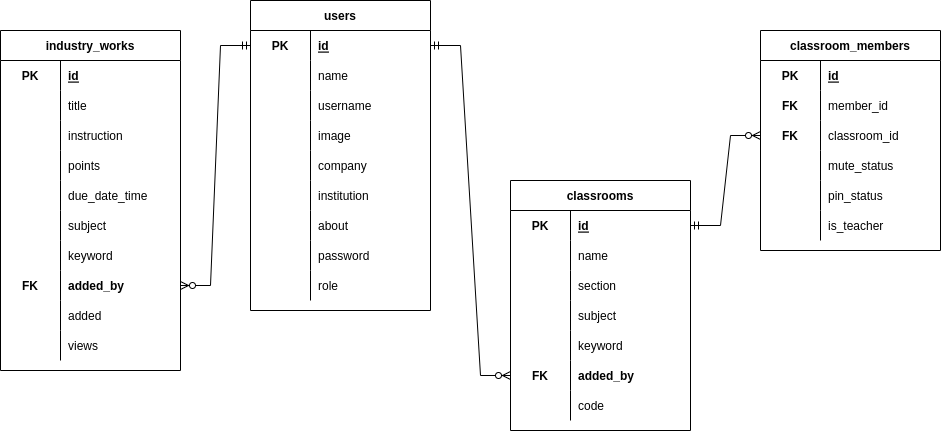
An entity relationship diagram, or ER Diagram or ERD, shows the connections between entity sets that are stored in a database. Our ER-model is divided into three main subject areas, as shown in Figure 4. 4, Figure 4. 5, and Figure 4. 6. The first diagram in Figure 4. 4 depicts the connections between the entities for *users*, *industry\_works*, and *classrooms*. Depending on the user role, a user can add multiple classrooms and industry tasks because the users entity has a one to many relationship with the *industry\_works* and *classrooms* entities. Since there is a one to many relationship between the *classroom\_members* and *classrooms* entity, multiple users can enrol in a classroom. The *classroom\_members* entity stores member id and classroom id as a foreign key. Attribute role determines whether a classroom member is its host or not.

Figure 4. 4 ER-Diagram

Figure 4. 5 begins by illustrating the connection between the classroom and the classroom posts. The *classrooms* entity has one to many relationship with *topics* and *classroom\_posts* entity, meaning that a classroom may have zero or more topics and classroom posts. A classroom post could be a *material*, *general\_post*, or *assignment* because the *classroom* entity has a one to one relationship with each of these entities. *Materials*, *general\_posts*, and *assignments* have a one to many relationship with their corresponding attachment entities, which means that these entities may have multiple attachments. Additionally, the *assignments* entity has a one-to-many relationship with the entities *assignment\_submissions* and *assignment\_comments*. *Assignment\_submissions* have a one-to-one relationship with *assignment submission\_grades*, and a one-to-many relationship with *assignment\_submission\_attachments*. It implies that a submission can have more than one attachment but only one grade.

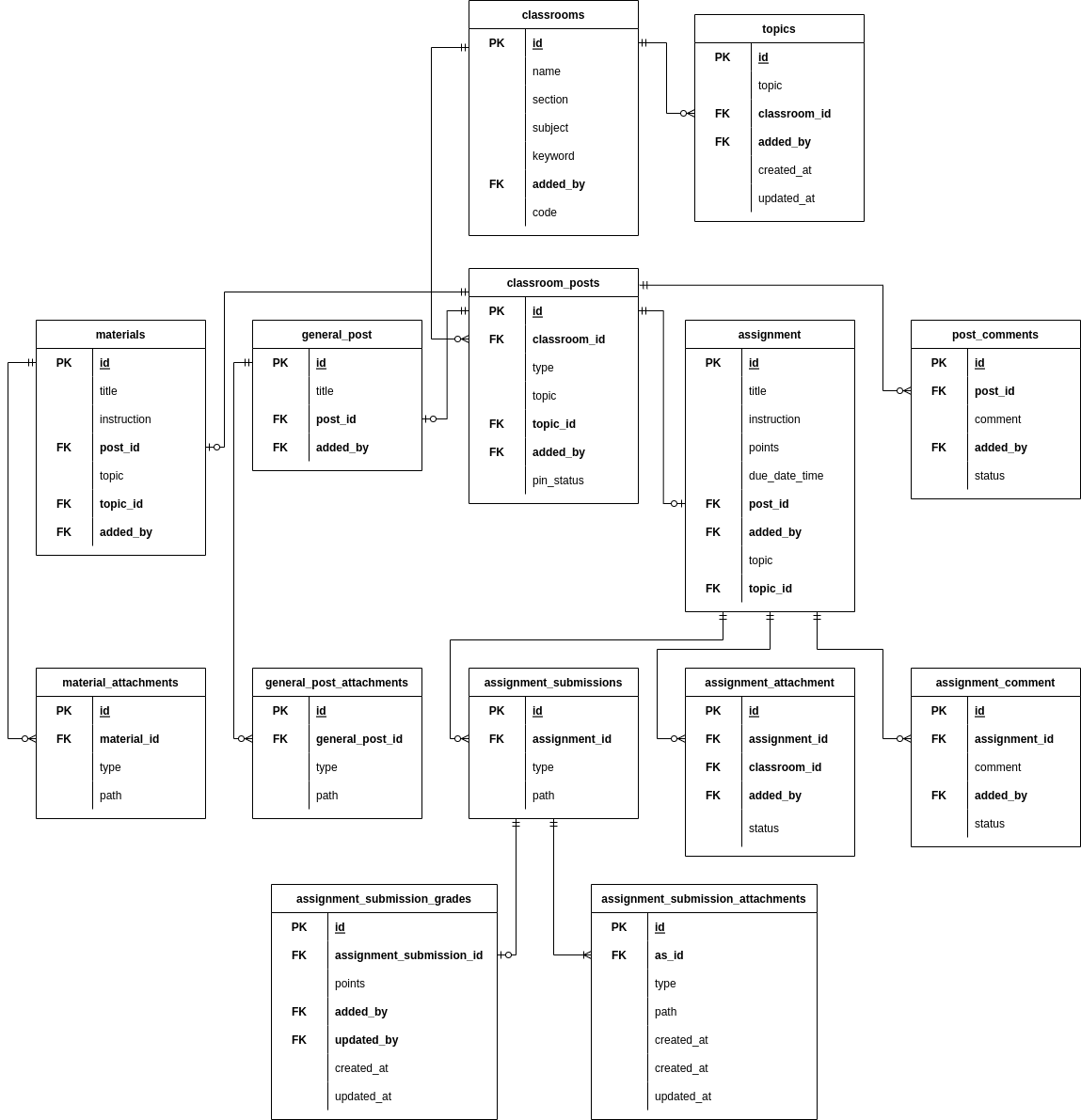


Figure 4. 5 ER-Diagram

Figure 4. 6 focuses on the subject industry work. As was previously mentioned, a user with the user role "industry" may have several industry tasks assigned for teachers to add to the classroom. *Industry\_works* has a **one to many** relationship with *industry\_work\_attachments* because industry work may have multiple attachments. Additionally, the *industry\_works* entity has a **one to many** relationship with the *classroom\_industry\_work* entity, allowing for the addition of an industry work in multiple classrooms. It's important to note that classroom industry work stores iw\_id and classroom\_id as a foreign key, enabling us to determine which industry work has been added to which classroom. Entity *classroom\_industry\_work* has a **one to many** relationship with *industry\_work\_submissions*, so an industry work may have multiple submissions. There may be multiple attachments but only one grade in a submission, so *industry\_work\_submissions* has a **one to many** relationship with *i\_w\_submission\_attachments* and a **one to one** relationship with *i\_w\_submission\_grades*.

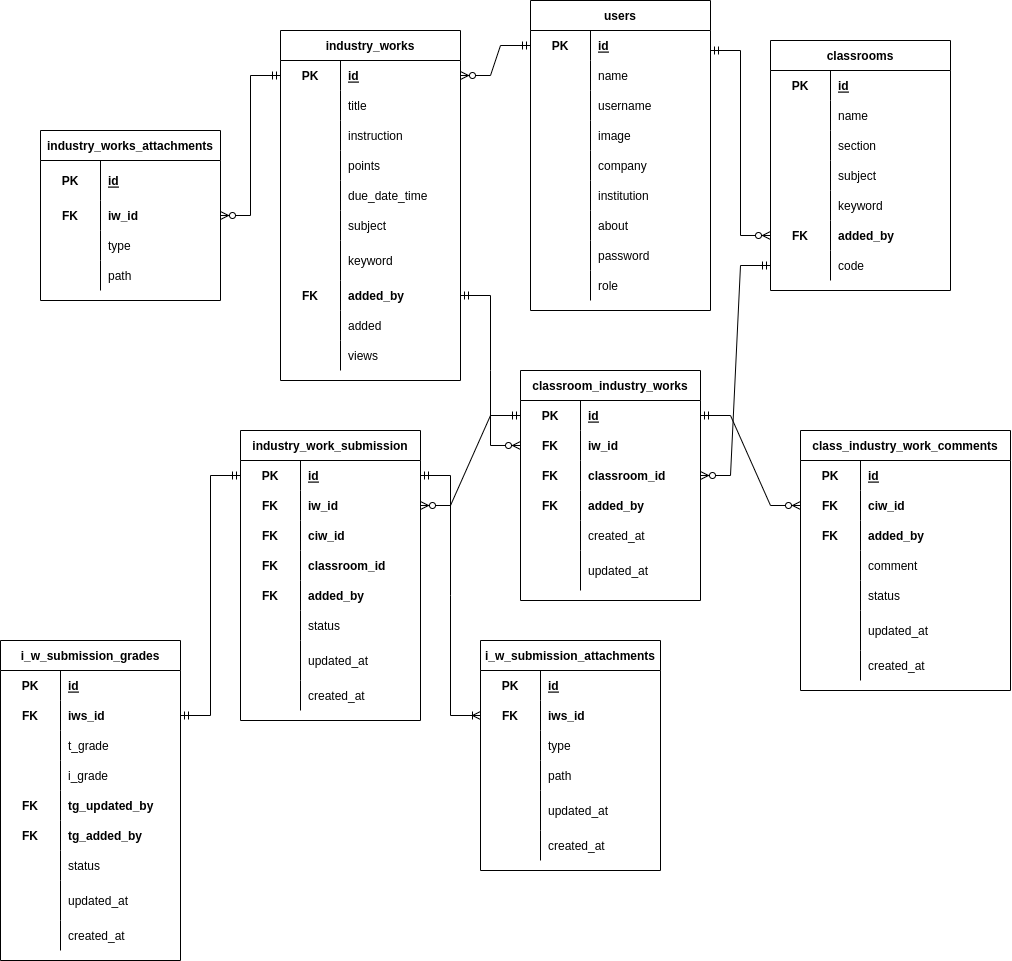


Figure 4. 6 ER-Diagram

**4.3 Use Case Diagram**

A use case explains how a user applies a system to reach a certain goal. It makes the details of how people interact with a system easier to understand.

Here, the system is used by the teacher, industry, and student as actors in academic activity. **Create classroom**, **post assignments**, **assigns grades**, and **adds jobs** are some top-level use cases for teacher. The teacher is able to create a new classroom by using the **create classroom** use case. Using the **add post** use case, a teacher can include an industry task into the classroom that has been added by the industry. Being able to **enter a classroom** and **submit assignments** and **industry jobs** would be another major use case for a teacher.

The student actor can join a classroom and submits an assignment using the **join classroom** and **submit assignment** case, just like the teacher can. A student may also want to view and submit industry jobs, which he can do by using the **view jobs** and **submit jo**bs use case.

Other major use cases, **post job** and **rate job submission** are specific to industry actor. The **post jobs** use case includes **add jobs** because an industry must post a job before it can be added in a classroom by a teacher.

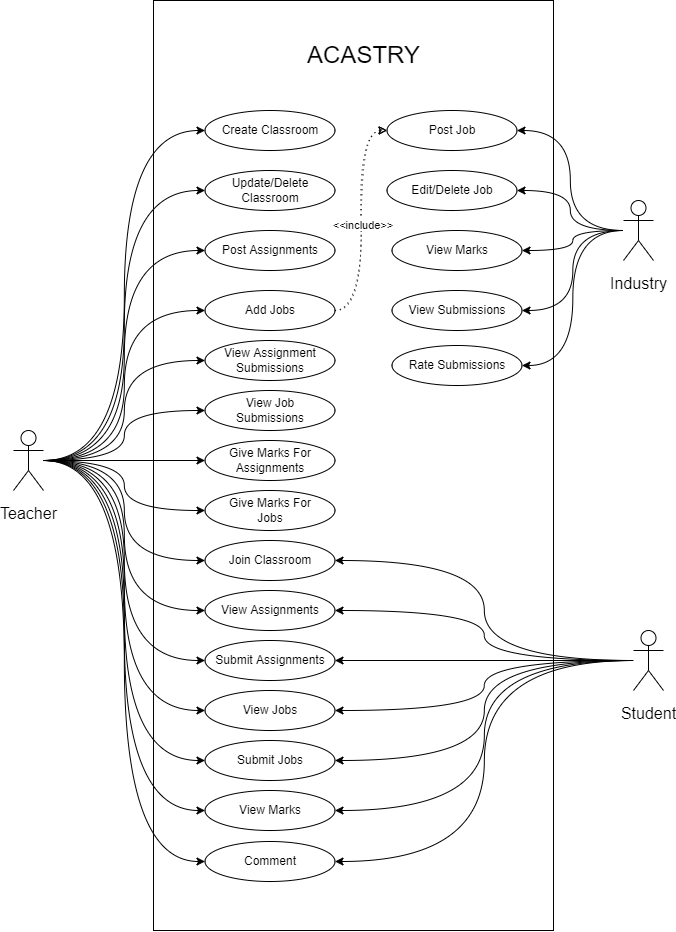
****

Figure 4. 7 Use Case Diagram

**4.4 System Flow Chart**

System flowcharts are a means to show how decisions are made to regulate events and how data flows in a system. The system begins when the users enter the website. Following user verification, the system will lead registered users to the login stage, while unregistered users are directed to the registration stage where they can enter their credentials or inputs. The system will then determine the user's role; if the user is a teacher, it will then permit them to create and join classrooms. When a teacher joins a class, they have the option in the following processes to view classroom posts or submit assignments and industry tasks. If a teacher creates a classroom, they can include industrial work or classroom posts in the classroom and they can assess those submissions afterwards.

On the other hand, a student can only enter a classroom. Therefore, a user with the "student" role who logs into the system can only join an existing classroom; he or she cannot create a new classroom. A student may also access the classroom posts and submit assignments and industry works. The following stage allows the student to view their grade if the assignment has been graded.

A user with the role "industry" will be redirected to the industry dashboard where industry can offer jobs/tasks for students. The following step allows an industry to review submitted work and provide grades for it. The system terminates when a user clicks the logout button to leave the system.

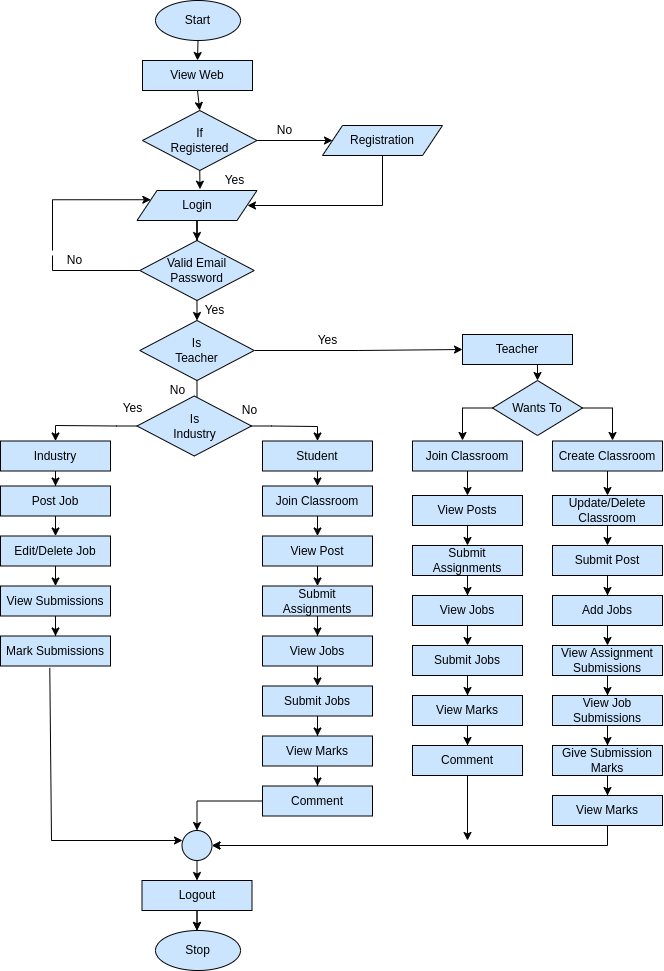


Figure 4. 8 System Flow Chart

**Chapter 5**

**Implementation Details**

**5.1.1 Login Page and Register Page**

Wherever the application is launched, a login page similar to **Figure 5. 1** appears. To create a new account, an unregistered user can go to the register page. To create a new account, a user must enter their name, username, email, and other necessary information.

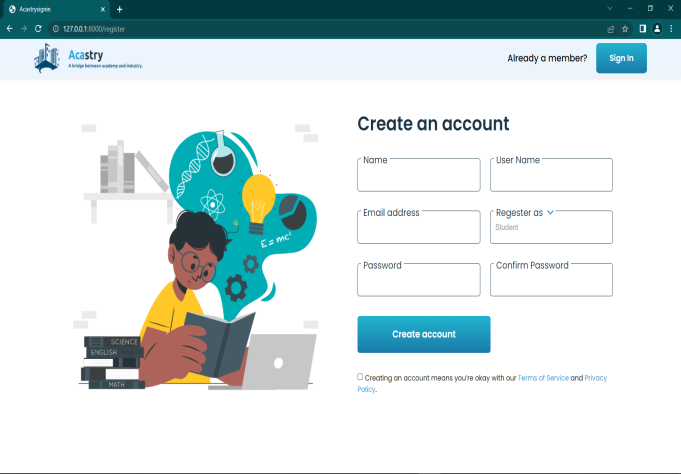


Figure 5. 1 Login Page and Registration Page

**5.1.2 Teacher, Student and Industry dashboard**

After successfully logging in, the user will be taken to the dashboard. A dashboard similar to **Figure 5. 2** will appear on the screen when a user with the roles of "student" or "teacher" logs into the system. A dashboard similar to **Figure 5. 3** will appear if an industry user logs into the system.

The dashboard offers a variety of options and data for the teacher and student to view. They can view the classroom they have joined or any assignments they might have. However, teachers can create new classrooms; students can only enter existing ones.

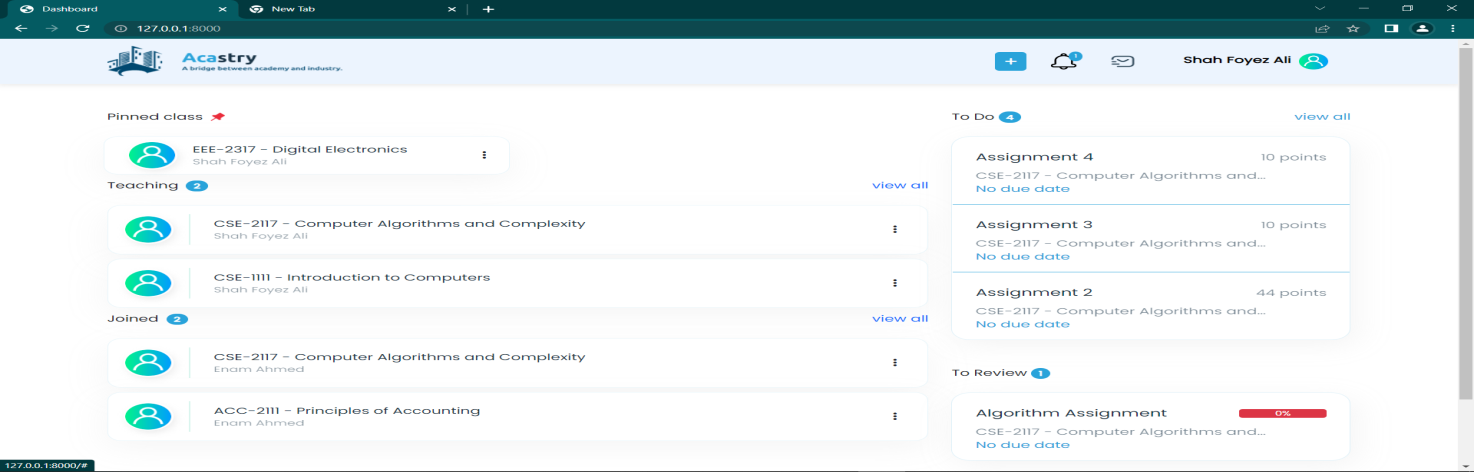
****

Figure 5. 2 Teacher and Student Dashboard

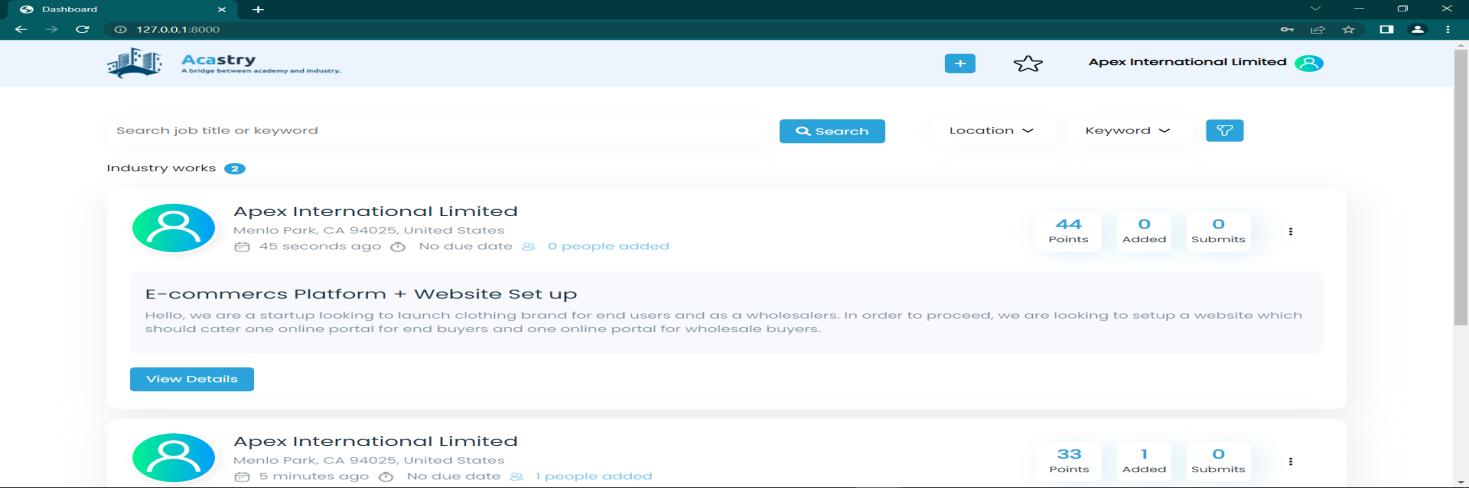
****

Figure 5. 3 Industry Dashboard

**5.1.3 Create or Join a Classroom**

A student user can only join a classroom; a teacher on the other hand can create and join classrooms in this system. A pop-up window where the user must enter the classroom title, subject, and other information appears when the teacher hits the "create classroom" button. A classroom code is automatically generated when a teacher creates a class. To join the class, a student must enter the classroom code in the join classroom popup depicted in **Figure 5. 5**

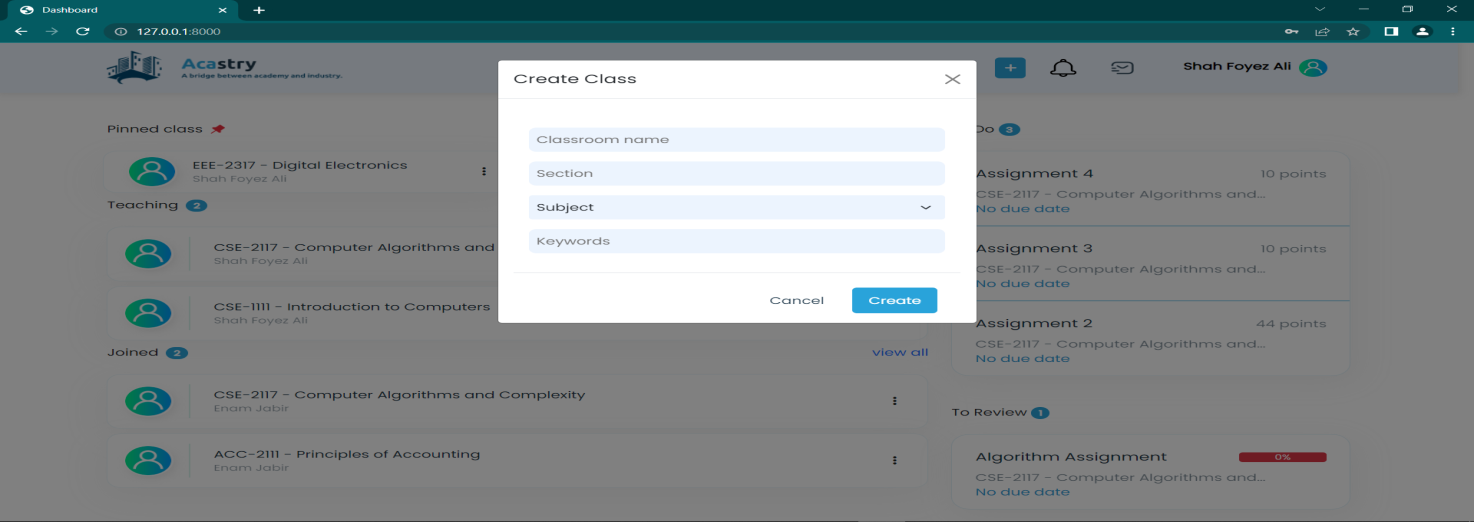
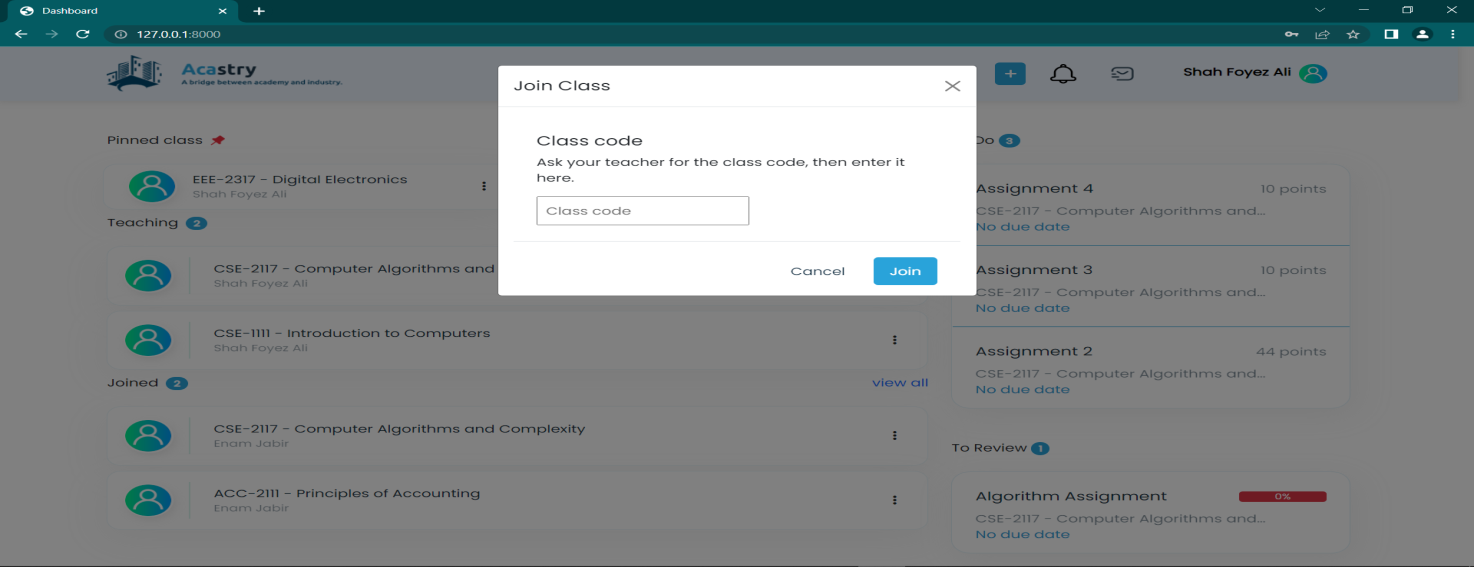


Figure 5. 4 Create Classroom Popup

**Figure 5. 5 Join Classroom Popup**

**5.1.4 Classroom Stream**

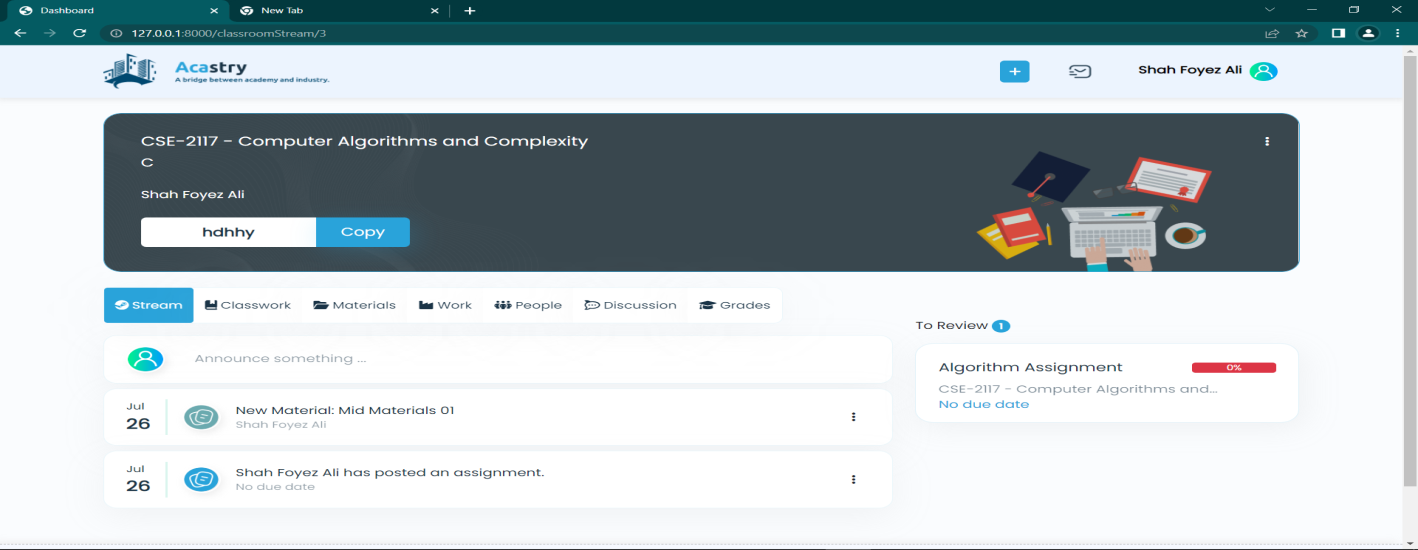
Students and teachers can both access the classroom stream. Every time a teacher posts something or adds new materials and assignments, it will appear on the classroom stream. Actions like edit or delete stream content are restricted, only the teacher has the access to modify stream content.

Figure 5. 6 Teacher Stream

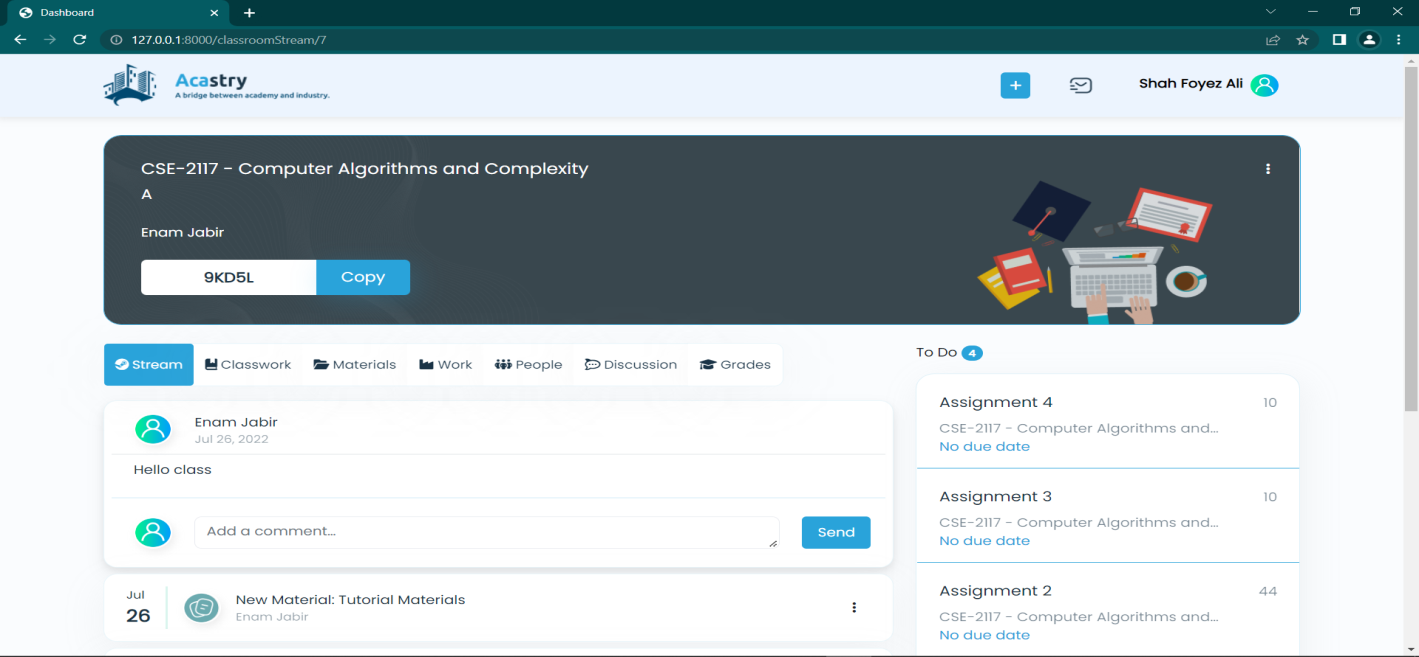
**5.1.5 Class work and Materials**

Figure 5. 7 Student Stream

The classroom's assignments and materials are shown on the class work and materials page. The classroom teacher is only permitted to add new assignments and materials, but both the teacher and students can view those classroom content. As seen in Figure 5. 8 and Figure 5. 9, a popup will appear when the teacher clicks the create button, allowing them to enter the information they need for the assignment or material.

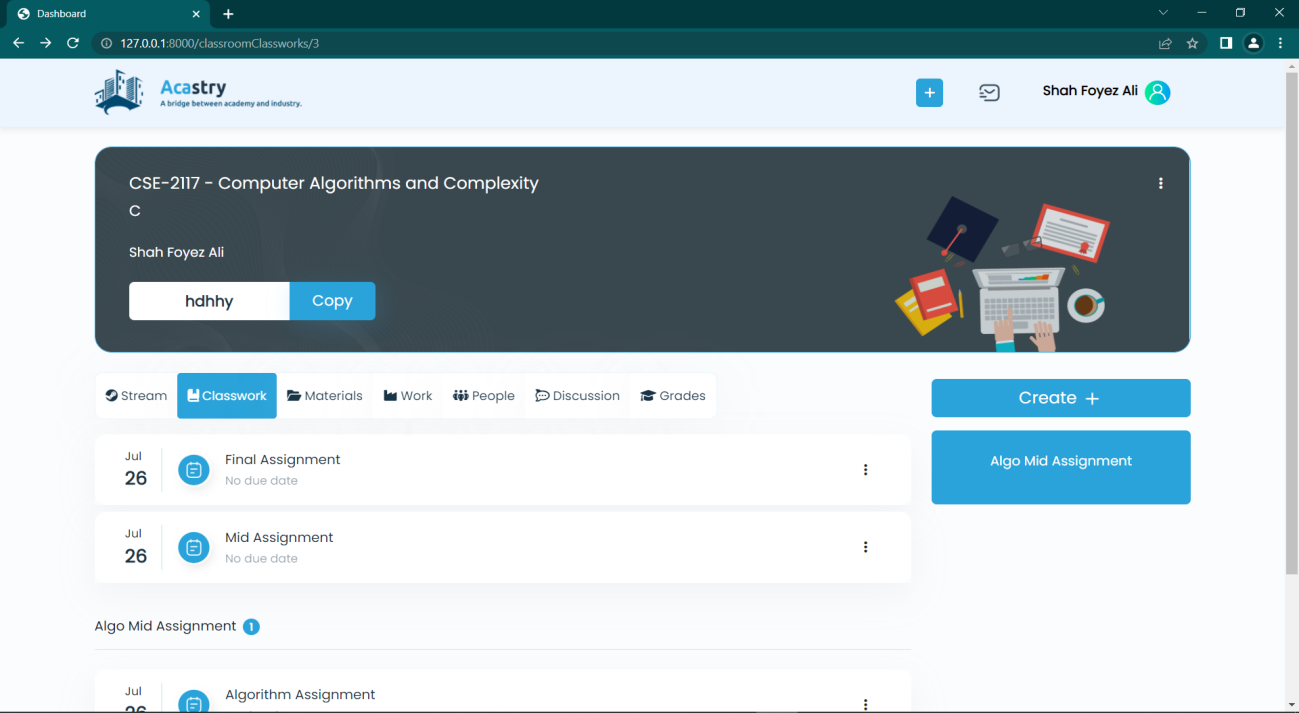
****

Figure 5. 8 Class work Page

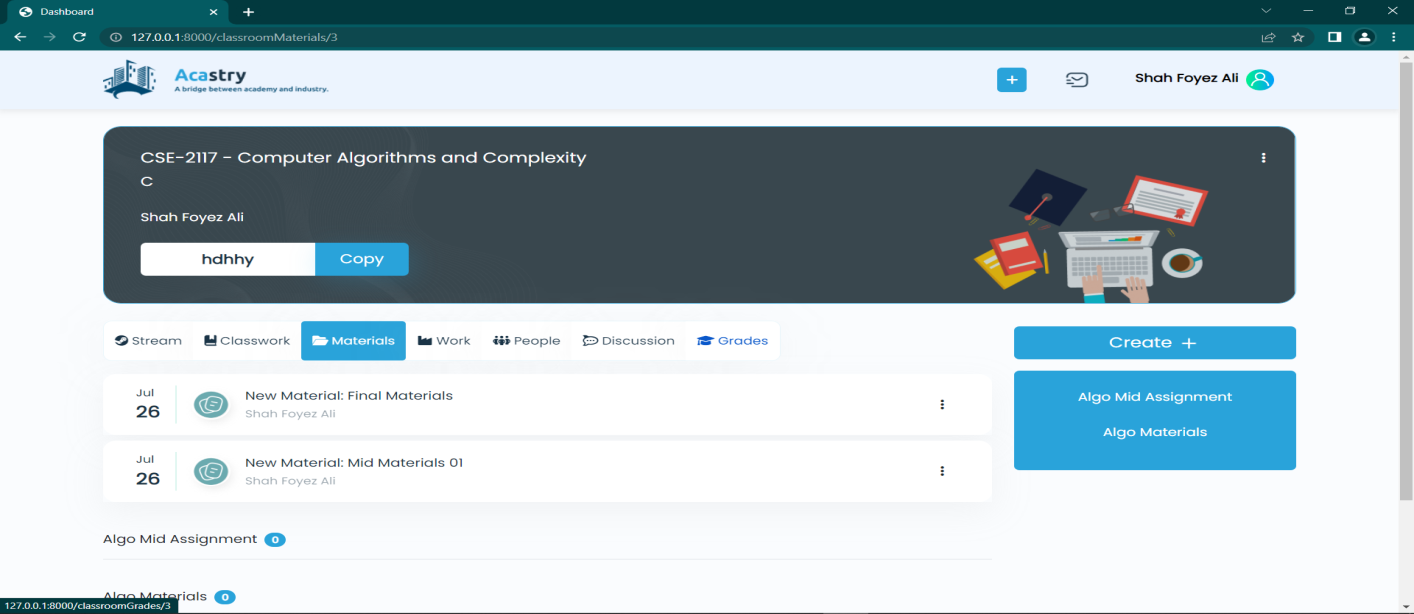
****

Figure 5. 9 Material Page

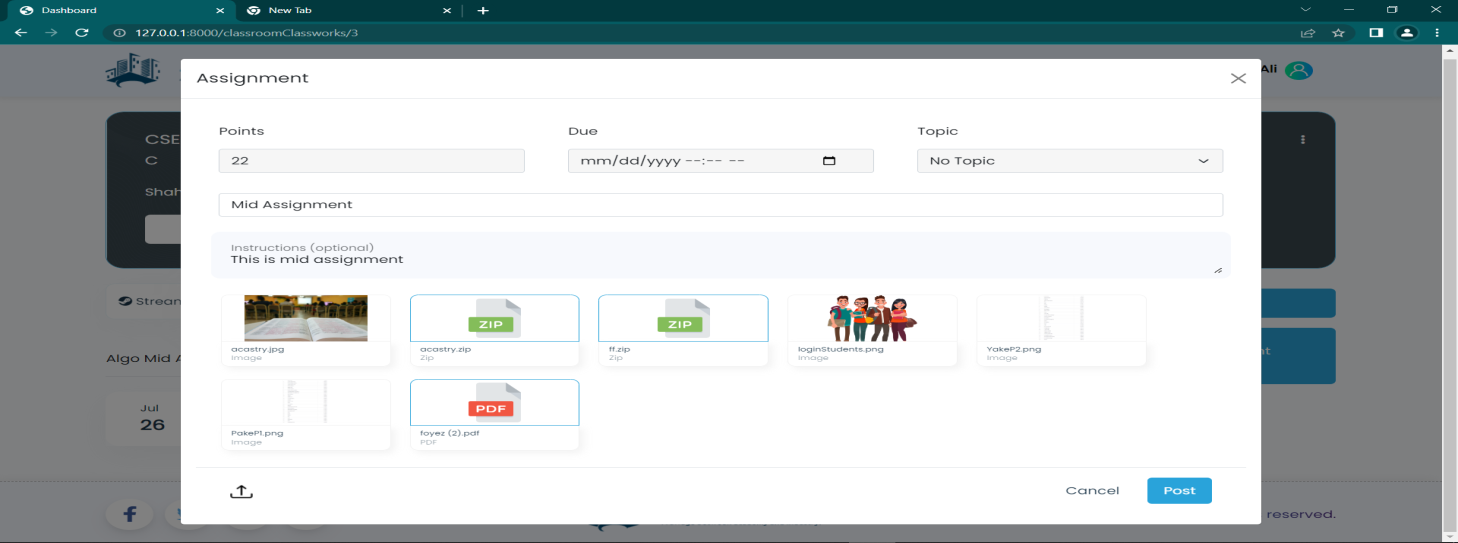
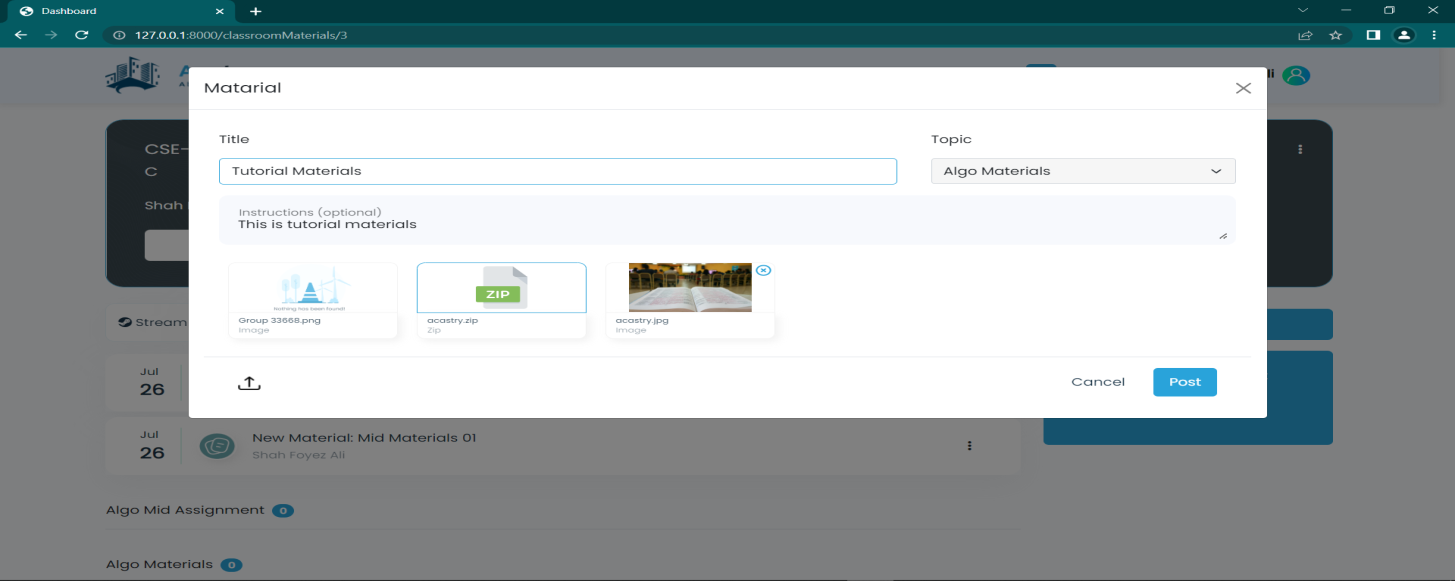
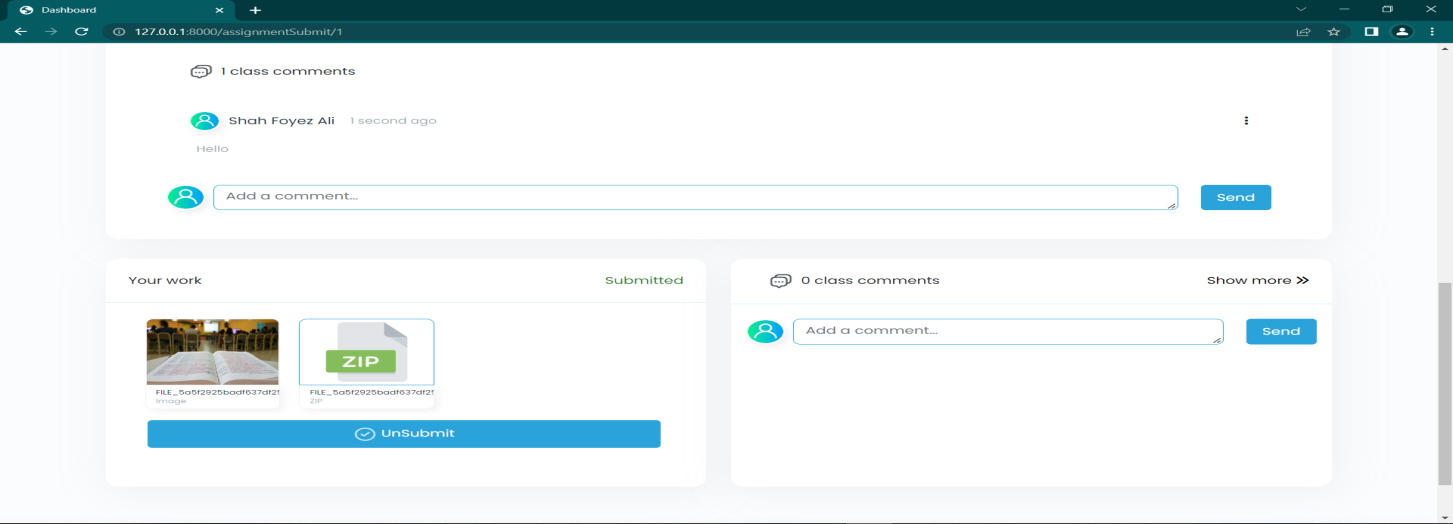
****

Figure 5. 10 Add assignment Popup

**Figure 5. 11 Add Material Popup**

**5.1.6 Assignment Submission and marking**

Students can see the instructions and upload their assignments on the assignment submit (**Figure 5.** 12) page. On the assignment submissions page, a teacher can view all of the assignment submissions (Figure 5. 13) and grade each one. Once their submission has been graded, a student can review their grade on the assignment submit (Figure 5. 12) page.



**Figure 5. 12 Assignment Submission Page for student**

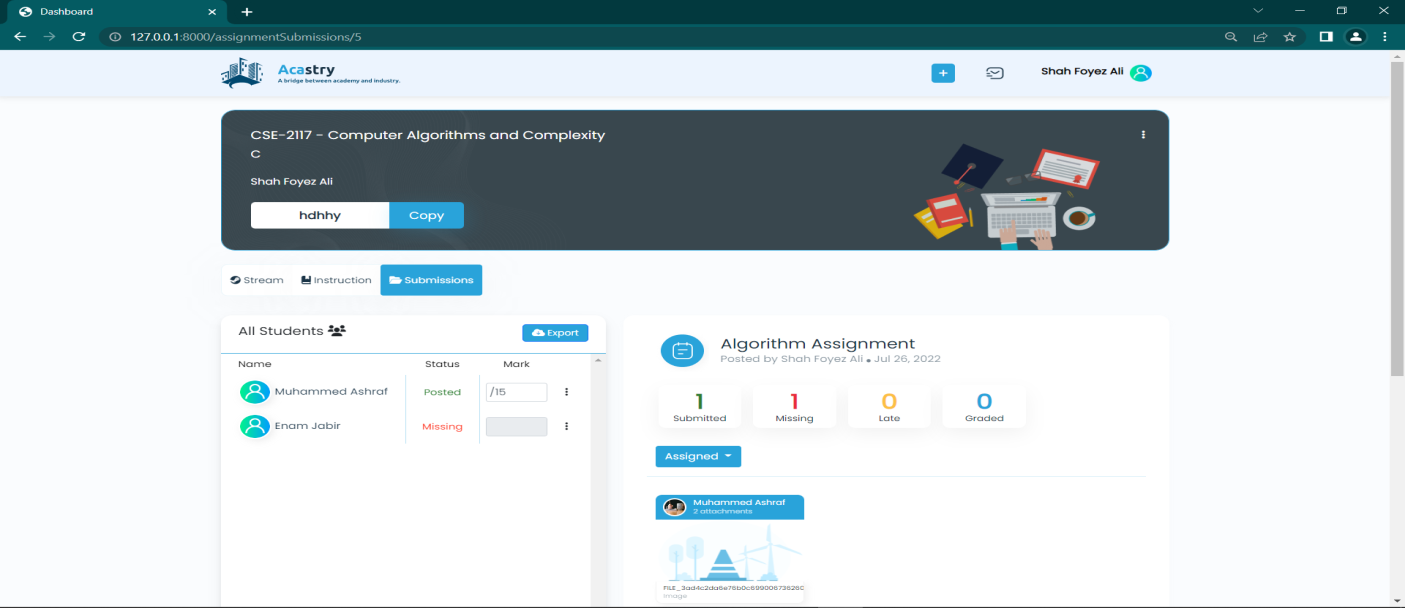
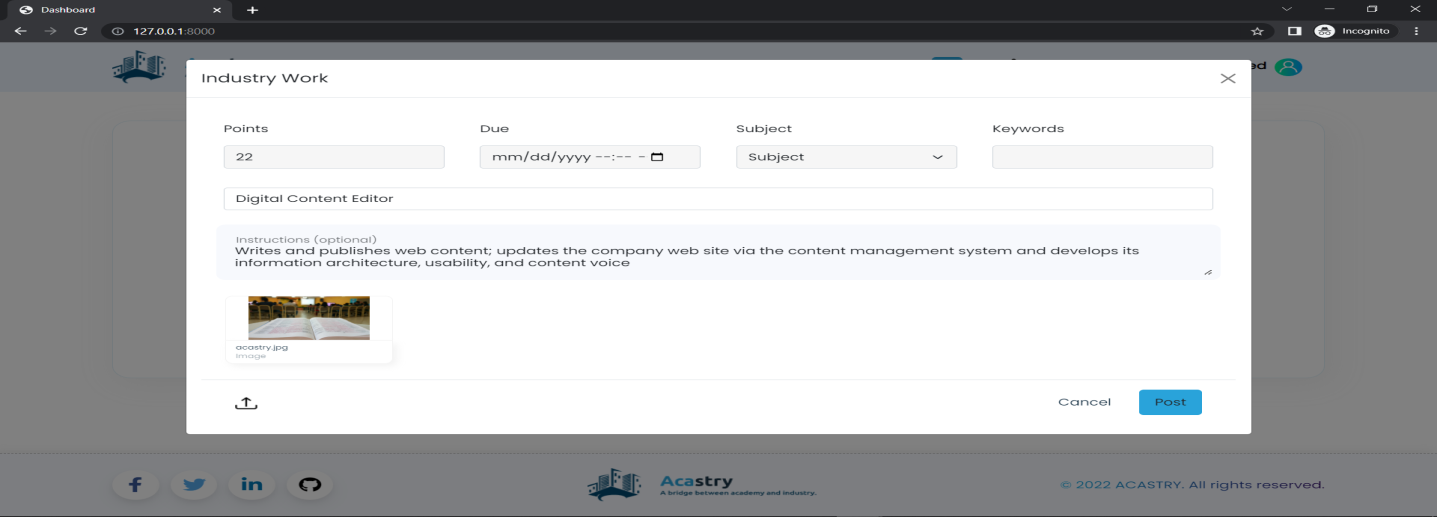
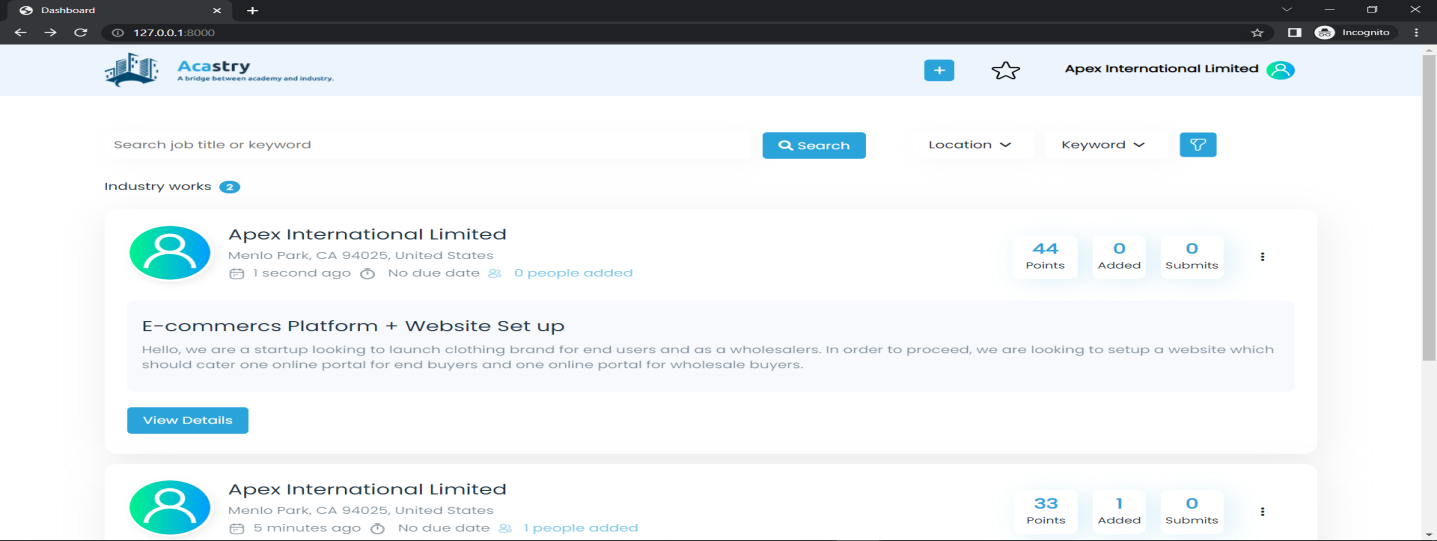


Figure 5. 13 Assignment Grade Page for Teacher

**5.1.7 Industry Work**

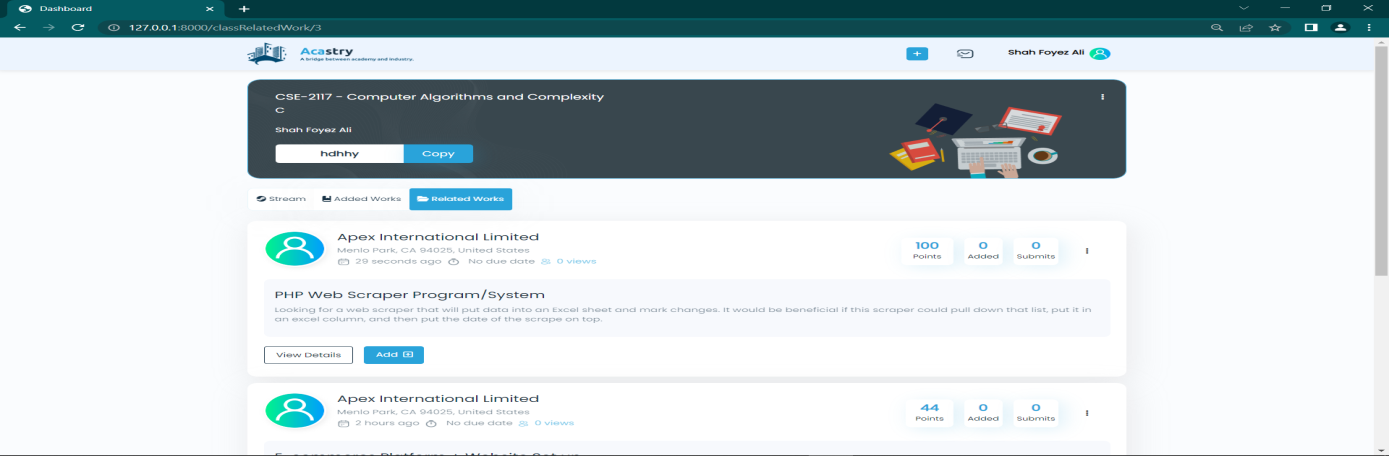
Figure 5. 14 shows the pop-up window that allows an industry user to post an industry work. An industry user must provide the required input to post an industry work. Once the work is posted, it will show up on the industry dashboard, as shown in Figure 5. 15.

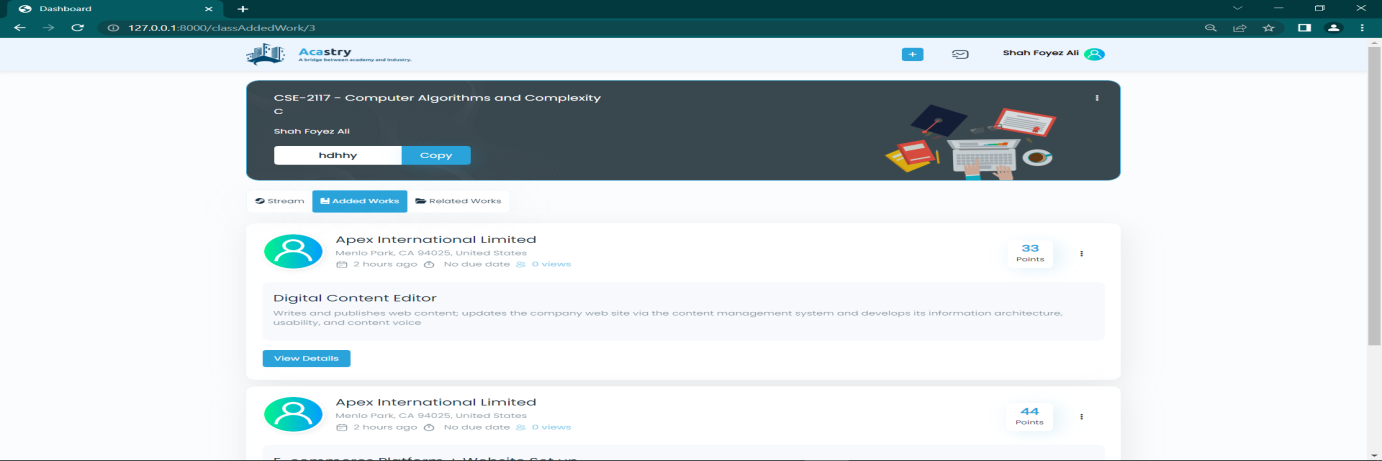
**Figure 5. 14 Add Industry Work Popup**

**Figure 5. 15 Industry Dashboard**

**5.1.8 Added work and related work**

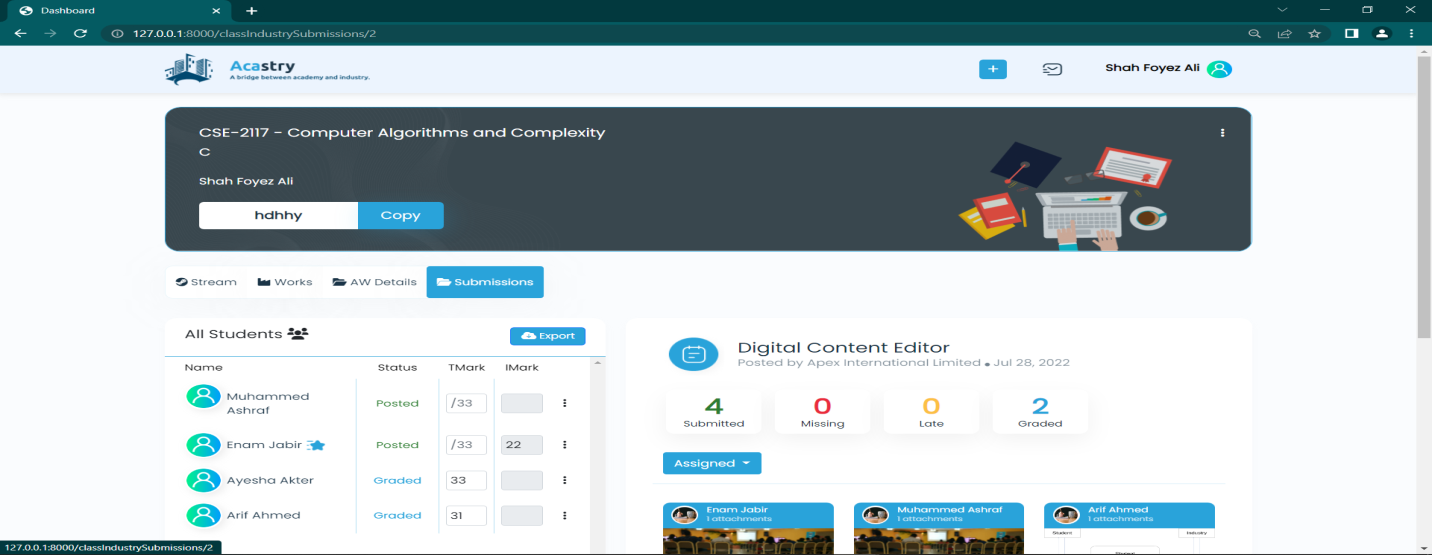
A teacher of a classroom might view his classroom-related industry works on the related work (Figure 5. 16) page. The teacher can review the work's details and add it to his class, allowing his students to submit the industry task. A task will appear in the added work page (Figure 5. 17) and be accessible for submission once it has been added to the classroom.

****

**Figure 5. 16 Related Work Page**

**Figure 5. 17 Added Work Page**

**5.1.9 Industry Work Submission and marking**

The submissions may be graded by both the teacher and an industry user; however a teacher may only grade student submissions from his or her own classroom. The grade that each party assigned to a task can be seen by both the teacher and industry practitioner.

**Figure 5. 18 Industry Work Submissions Page for Teacher**

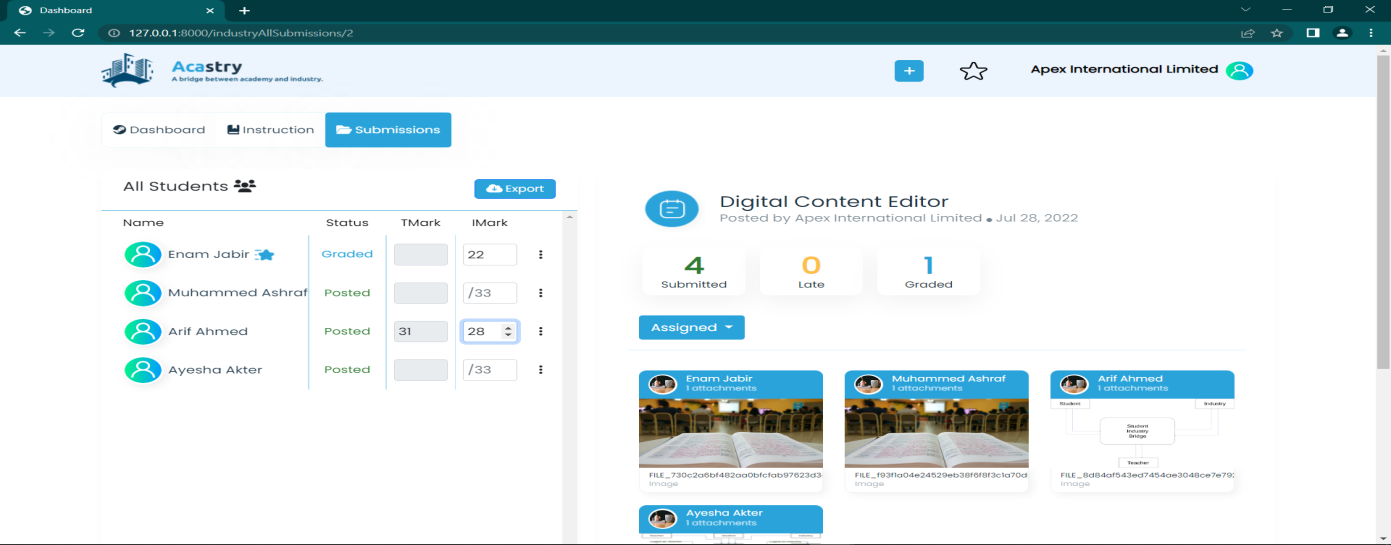
****

Figure 5. 19 Industry Work Submissions Page for Industry

**Chapter 6**

Results and Discussion

* 1. **Datasets**

We put our suggested keyword extraction system to the test on two different document collections to gauge its effectiveness. The collections include texts with varying text lengths, and we have made use of well-known databases like Inspec [21] and SemEval2010 [27]. The details of each dataset are as follows:

The 244 complete scientific papers obtained for SemEval2010 [48] from the ACM Digital Library are one of the most well-known datasets previously used for keyword extraction evaluation. Each research paper consists of a set of keywords selected by the individual author along with a sizable group of keywords chosen by professional editors; the input text might not necessarily contain keywords from both of the given sets. This dataset ranges in length from 6 to 8 pages spanning over four different fields of computer science research which are as follows [48]:

* Social and behavioral sciences (economics)
* Distributed systems, information search, and retrieval
* Distributed artificial intelligence (multiagent systems)

Moreover, the specific version of the dataset used for evaluation refers to the works of (Boudin et al., 2016), developed by their team to incorporate four refined standards for the overall processing of documents. It included manual extraction of the PDF files, processing the input into suitable columns, and choosing the most content-rich documents in the process.

The Inspec [24] dataset includes 2000 abstracts of computer science journal articles gathered between 1998 and 2002. It focuses on the following disciplines: Computers and Control; and Information Technology. Each article has its respective title along with the keywords obtained from Inspec's database, and a designated indexer assigned two sets of keywords to each abstract: a collection of controlled terms, or words strictly adhering to the Inspec thesaurus, and a group of uncontrolled terms that may contain any relevant word. At the same time, it may not be necessary that the abstracts would consist of terms from both sets of controlled and uncontrolled words.

* 1. **Performance Metrics**

The most important question that arises is how good our model is at supporting the system. So, evaluating the proposed model is essential to delineate how effectively our system can extract the keywords. This requires specific performance metrics to demonstrate the evaluation.

Beforehand, we need to understand the underlying terminologies:

**True Positives (TP)**: These are the correctly predicted positive values stating that the value of the actual class and the value of the predicted class are both true. In this case, the extracted keywords and the golden keywords are the same.

**True Negatives (TN)**: These are the correctly predicted negative values inferring that the value of the actual class and value of the predicted class both are false. This means the extracted keywords and the golden keywords do not match

**False positives (FP)** and **False Negatives (FN)** happens when the actual class contradicts the predicted class.

We will elaborate on the following metrics taken into consideration below.

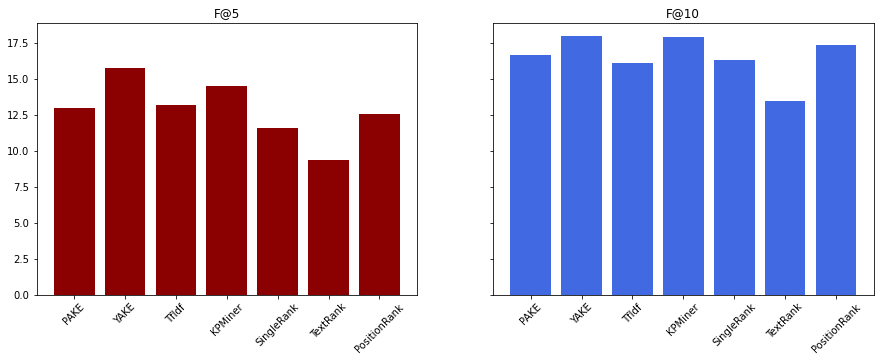
**Precision** - Precision is the ratio of correctly predicted positive observations to the total predicted positive observations. High precision relates to the low false positive rate. In other words, it measures how many positive predictions are correct (true positives).

**Recall**(Sensitivity) - Recall is the ratio of correctly predicted positive observations to all observations in the actual class - True.

**F1 score** - F1-Score is a measure combining precision and recall. It is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. F1 is usually more helpful than accuracy when there is an uneven class distribution. F1-score becomes high only when both precision and recall are high.

* 1. **Comparison Analysis**

The distribution of the top-N candidate keywords is conventionally based on precisely matching the golden keywords with the extracted ones. In some contexts of literature review, keywords that are close enough semantically are also considered as part of the standard keywords of the given corpus. Since our task involves the Inspec and the SemEval2010 dataset, we pursue the traditional approach of matching the keywords exactly. We determine the precision, recall, and F-score by comparing the provided presumed key phrases in the dataset. The performance of each system is evaluated by examining the output as the top-5 and top-10 candidate keywords. A summary of the results is shown in **Table 4**, collates PAKE's usefulness with that of the baselines on each dataset in percentage terms.

**Figure 6.3. 1 F-scores on the SemEval2010 Dataset**

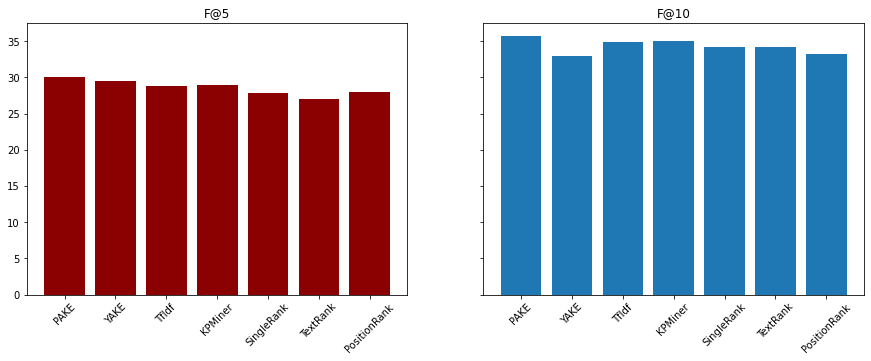


Figure 6.3. 2 F-scores on the Inspec Dataset

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Dataset** | **Total Documents** | **PAKE** | | **YAKE** | | **TF-IDF** | | **KPminer** | | **SingleRank** | | **TextRank** | | **PositionRank** | |
| **F@5** | **F@10** | **F@5** | **F@10** | **F@5** | **F@10** | **F@5** | **F@10** | **F@5** | **F@10** | **F@5** | **F@10** | **F@5** | **F@10** |
| Inspec | 244 | 30.13 | 35.75 | 29.54 | 32.94 | 28.75 | 34.94 | 28.93 | 35.07 | 27.80 | 34.18 | 27.03 | 34.20 | 28.04 | 33.20 |
| SemEval2010 | 2000 | 12.93 | 16.63 | 15.73 | 17.97 | 13.20 | 16.08 | 14.48 | 17.90 | 11.55 | 16.29 | 9.32 | 13.42 | 12.54 | 17.32 |

**Table 4 Comparison of F-scores on the datasets**

It is evident to corroborate that the overall efficacy of our proposed keyword extraction is substandard for the following reasons: (1) The statistical methods do not filter out the corpus on the basis of lexical terms as compared to PAKE and other graph-based models, thereby resulting in biased results **Figure 6.3. 1**. (2) The certain constraints faced in any NLP task along with the lack of a customised dataset.

We conclude this analysis by suggesting that f-scores show that overall, the results increase and are higher than the graph-based models that employ POS taggers. Furthermore, a slight contrast arises when compared to PositionRank, which marginally provides favorable results. Finally, regarding the statistical methods, the system's output of f-scores exhibits considerably lower scores than statistical baselines, especially YAKE! and KPminer for the SemEval2010 dataset. However, for F@10, the improvement in the scores indicates better precision and recall. Remarkably, the performance boosts when using the Inspec Dataset and PAKE subsequently has higher F-scores in contrast to the mentioned baselines. Therefore, we can infer that PAKE is a plausible substitute for state-of-the-art graph-based algorithms to extract keywords from a given corpus concerning the usage of POS taggers and can be significantly enhanced to match the performance concerning the statistical baselines.

* 1. **Discussion**

The substantial performance of our proposed system spans over the chosen number of candidate keywords and the varying categories of author-assigned and professional indexer-assigned keywords. The evaluation demonstrates the relevant achievement of our proposed keyword extraction system compared to some of the existing graph-based alternatives and is conditionally at par with the statistical baselines.

We can assume that obtaining keywords automatically depends on the particular study's aims. Hence, it can be subjective from the perspective of a researcher. Undoubtedly, the system can continually be improved for effectiveness as well as better performance. However, irrespective of using POS tagging, there is still a need for an evaluation in terms of the semantic understanding of the candidate keywords. This will eventually lead to a better approach to generating keywords with improved results in evaluation metrics.

**Chapter 7**

Conclusion and Future Work

**7.1 Conclusion**

The proposed system attempts to mitigate the differences between academia and the industry with the intent of successful collaboration through our AI-influenced system. The web application is aided by a POS tagger Augmented Keyword Extraction system that eventually works to provide better recommendations to the respective user. The potential benefits are highlighted throughout the report, and we believe the system will benefit both students and industry practitioners. If we consider from the student's perspective, the system will help the student gain relevant industry experience by understanding the kind of projects available at the industry level. Thus, refining the profile with adequate skills will help the student enter the job market. From the other perspective, it is a better way to hire talented students or possibly outsource students to complete existing projects. This could potentially be a way for industries to include the component of corporate social responsibility and help minimize the growing gap.

**7.2 Future Work**

We have only been able to carry out hands-on work with ideas in a few areas and some preliminary work on evaluating the proposed modules for the system. Although we feel that substantial progress has been made towards the overall goal of defining a general framework for the design of evaluation methodologies, we are aware that much remains to be done. In particular, we intend to extend the keyword extraction system with plausible semantic meaning attached to it and build a supervised model that will help generate relevant skills specific to the skills required for the various jobs in the IT sector. It would require an elaborate and exhaustive dataset of keywords depicting the skills and thus is a future endeavor. Moreover, we intend to compare various models using different recommendation techniques to evaluate the overall performance to decide the best suitable approach for our existing system. This may involve using a content-based method aided with machine learning algorithms such as Naive Bayes etc. It may possibly extend to using collaborative filtering to connect the users at a community level or increase the overall efficiency by using hybrid recommenders for the system. In the future, the amalgamation of all the scenarios mentioned earlier can elevate the performance of our AI-based web app and eventually achieve the goal of reducing the gap to a large extent.

**References**

[1] [World Bank. 2019. Bangladesh Development Update, October 2019 : Tertiary Education and Job Skills. World Bank, Dhaka. © World Bank.   
https://openknowledge.worldbank.org/handle/10986/32533 License: CC BY 3.0 IGO](file:///C:\Users\Staff%20Asia%20-%20PC\Desktop\Reports\World%20Bank.%202019.%20Bangladesh%20Development%20Update,%20October%202019%20:%20Tertiary%20Education%20and%20Job%20Skills.%20World%20Bank,%20Dhaka.%20©%20World%20Bank.%20%20https:\openknowledge.worldbank.org\handle\10986\32533%20License:%20CC%20BY%203.0%20IGO)[.](https://openknowledge.worldbank.org/handle/10986/32533)

[2][Khatun, F., Saadat, S. Y., Ashraf, K., & Taki, M. A. T. (2022). Skills Gap and Youth Employment in Bangladesh: An Exploratory Analysis. Dhaka: Centre for Policy Dialogue (CPD) and Friedrich-Ebert-Stiftung (FES) Bangladesh](https://cpd.org.bd/publication/skills-gap-and-youth-employment-in-bangladesh/)

[3] N. Firoozeh, A. Nazarenko, F. Alizon and B. Daille, "Keyword extraction: Issues and methods", Natural Language Engineering, vol. 26, no. 3, pp. 259-291, 2019. Available: 10.1017/s1351324919000457.

[4] S. Beliga, “Keyword extraction: a review of methods and approaches,”  
University of Rijeka, Department of Informatics, Rijeka, vol. 1, no. 9,  
2014.

[5] P. D. Turney, “Learning algorithms for keyphrase extraction,” Informa-  
tion retrieval, vol. 2, no. 4, pp. 303–336, 2000.

[6] T. Nguyen and M. Kan, "Keyphrase Extraction in Scientific Publications", Asian Digital Libraries. Looking Back 10 Years and Forging New Frontiers, pp. 317-326. Available: 10.1007/978-3-540-77094-7\_41.

[7] I. Witten, G. Paynter, E. Frank, C. Gutwin and C. Nevill-Manning, "KEA", Design and Usability of Digital Libraries, pp. 129-152, 2005. Available: 10.4018/978-1-59140-441-5.ch008.

[8] O. Medelyan and I. Witten, "Thesaurus based automatic keyphrase indexing", Proceedings of the 6th ACM/IEEE-CS joint conference on Digital libraries - JCDL '06, 2006. Available: 10.1145/1141753.1141819.

[9] E. Papagiannopoulou and G. Tsoumakas, “A review of keyphrase extraction,” Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 10, no. 2. Wiley-Blackwell, Mar. 01, 2020. doi: 10.1002/widm.1339.

[10] R. Campos, V. Mangaravite, A. Pasquali, A. Jorge, C. Nunes, and A. Jatowt, “YAKE! Keyword extraction from single documents using multiple local features,” *Information Sciences*, vol. 509, pp. 257–289, Jan. 2020, doi: 10.1016/j.ins.2019.09.013.

[11] K. S. Hasan and V. Ng, “Automatic keyphrase extraction: A survey of the state of the art,” in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2014, pp. 1262–1273

[12] G. Salton, A. Wong, and C. S. Yang, “A Vector Space Model for Automatic Indexing,” *Commun ACM*, vol. 18, no. 11, pp. 613–620, Nov. 1975, doi: 10.1145/361219.361220.

[13] Florescu Corina and Caragea Cornelia, “A New Scheme for Scoring Phrases in Unsupervised Keyphrase Extraction,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 10193 LNCS. Springer Verlag, pp. V–VI, 2017. doi: 10.1007/978-3-319-56608-5.

[14] S. R. El-Beltagy and A. Rafea, “KP-Miner: A keyphrase extraction system for English and Arabic documents,” *Information Systems*, vol. 34, no. 1, pp. 132–144, Mar. 2009, doi: 10.1016/j.is.2008.05.002.

[15] Z. Liu, P. Li, Y. Zheng, and M. Sun, “Clustering to find exemplar terms  
for keyphrase extraction,” in Proceedings of the 2009 conference on  
empirical methods in natural language processing, 2009, pp. 257–266.

[16] S. Rose, D. Engel, N. Cramer, and W. Cowley, “Automatic keyword  
extraction from individual documents,” Text mining: applications and  
theory, vol. 1, no. 1-20, pp. 10–1002, 2010.

[17] M. Won, B. Martins, and F. Raimundo, “EasyChair Preprint Automatic extraction of relevant keyphrases for the study of issue competition Automatic extraction of relevant keyphrases for the study of issue competition,” 2019. [Online]. Available: <https://manifesto-project.wzb.eu>

[18] T. Tomokiyo and M. Hurst, “A language model approach to keyphrase extraction,” in Proceedings of the ACL 2003 workshop on Multiword expressions: analysis, acquisition and treatment, 2003, pp. 33–40.

[19] Y.-H. Woo, D.-H. Nam, T.-S. Hur, Y.-B. Park, W. Huh, Y.-S. Woo, and H.-K. Min, “Automated keyword extraction using category correlation of data,” in International Conference on Computational Science and Its Applications. Springer, 2006, pp. 224–230.

[20] R. Mihalcea and P. Tarau, “Textrank: Bringing order into text,” in  
Proceedings of the 2004 conference on empirical methods in natural  
language processing, 2004, pp. 404–411.

[21] X. Wan and J. Xiao, “Single document keyphrase extraction using neighborhood knowledge.” in AAAI, vol. 8, 2008, pp. 855–860.

[22] S. Danesh, T. Sumner, and J. H. Martin, “SGRank: Combining statistical and graphical methods to improve the state of the art in unsupervised keyphrase extraction,” in Proceedings of the fourth joint conference on lexical and computational semantics, 2015, pp. 117–126.

[23] C. Florescu and C. Caragea, “PositionRank: An unsupervised approach to keyphrase extraction from scholarly documents,” in *ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers)*, 2017, vol. 1, pp. 1105–1115. doi: 10.18653/v1/P17-1102.

[24] A. Hulth, “Improved automatic keyword extraction given more linguis-  
tic knowledge,” in Proceedings of the 2003 conference on Empirical  
methods in natural language processing, 2003, pp. 216–223

[25] O. Medelyan, E. Frank, and I. H. Witten, “Human-competitive tagging using automatic keyphrase extraction.” Association for Computational Linguistics, 2009

[26] S. D. Gollapalli and C. Caragea, “Extracting keyphrases from research  
papers using citation networks,” in Proceedings of the AAAI conference  
on artificial intelligence, vol. 28, no. 1, 2014.

[27] S. D. Gollapalli, X.-L. Li, and P. Yang, “Incorporating expert knowledge into keyphrase extraction,” in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 31, no. 1, 2017.

[28] E. Papagiannopoulou and G. Tsoumakas, “Local word vectors guiding keyphrase extraction,” *Information Processing and Management*, vol. 54, no. 6, pp. 888–902, Nov. 2018, doi: 10.1016/j.ipm.2018.06.004.

[29] "Nvidia Jetson emmc full encryption with secure boot | Linux | Security | Encryption | Hardware Security Module | Freelancer", Freelancer.com, 2022. [Online]. Available: https://www.freelancer.com/projects/linux/nvidia-jetson-emmc-full-encryption. [Accessed: 20- Jun- 2022].

[30] M. Rahman and N. Abdullah, "A Personalized Group-Based Recommendation Approach for Web Search in E-Learning", *IEEE Access*, vol. 6, pp. 34166-34178, 2018. Available: 10.1109/access.2018.2850376

[31] A. Zehtab-Salmasi, M.-R. Feizi-Derakhshi, and M.-A. Balafar, “Frake:Fusional real-time automatic keyword extraction,” arXiv preprintarXiv:2104.04830, 2021.

[32] Q. Zhang, J. Lu and Y. Jin, "Artificial intelligence in recommender systems", Complex &amp; Intelligent Systems, vol. 7, no. 1, pp. 439-457, 2020. Available: 10.1007/s40747-020-00212-w.

[33] S. Khanal, P. Prasad, A. Alsadoon and A. Maag, "A systematic review: machine learning based recommendation systems for e-learning", Education and Information Technologies, vol. 25, no. 4, pp. 2635-2664, 2019. Available: 10.1007/s10639-019-10063-9.

[34] A. Felfernig, M. Jeran, G. Ninaus, F. Reinfrank, S. Reiterer and M. Stettinger, "Basic Approaches in Recommendation Systems", Recommendation Systems in Software Engineering, pp. 15-37, 2013. Available: 10.1007/978-3-642-45135-5\_2.

[35] B. Mobasher, “Data mining for web personalization,” in The adaptive web. Springer, 2007, pp. 90–135.

[36] F. Ricci, L. Rokach and B. Shapira, "Introduction to Recommender Systems Handbook", Recommender Systems Handbook, pp. 1-35, 2010. Available: 10.1007/978-0-387-85820-3\_1.

[37] S. T. Al-Otaibi, "A survey of job recommender systems", International Journal of the Physical Sciences, vol. 7, no. 29, 2012. Available: 10.5897/ijps12.482.

[38] S. Khanal, P. Prasad, A. Alsadoon and A. Maag, "A systematic review: machine learning based recommendation systems for e-learning", Education and Information Technologies, vol. 25, no. 4, pp. 2635-2664, 2019. Available: 10.1007/s10639-019-10063-9.

[39] N. Almalis, G. Tsihrintzis, N. Karagiannis and A. Strati, "FoDRA — A new content-based job recommendation algorithm for job seeking and recruiting", 2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA), 2015. Available: 10.1109/iisa.2015.7388018.

[40] G. Adomavicius and A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions", IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 6, pp. 734-749, 2005. Available: 10.1109/tkde.2005.99.

[41] P. Lops, M. de Gemmis and G. Semeraro, "Content-based Recommender Systems: State of the Art and Trends", Recommender Systems Handbook, pp. 73-105, 2010. Available: 10.1007/978-0-387-85820-3\_3.

[42] P. B.Thorat, R. M. Goudar and S. Barve, "Survey on Collaborative Filtering, Content-based Filtering and Hybrid Recommendation System", International Journal of Computer Applications, vol. 110, no. 4, pp. 31-36, 2015. Available: 10.5120/19308-0760.

[43] F. Sebastiani, "Machine learning in automated text categorization", ACM Computing Surveys, vol. 34, no. 1, pp. 1-47, 2002. Available: 10.1145/505282.505283.

[44] J. Herlocker, J. Konstan, L. Terveen and J. Riedl, "Evaluating collaborative filtering recommender systems", ACM Transactions on Information Systems, vol. 22, no. 1, pp. 5-53, 2004. Available: 10.1145/963770.963772.

[45] Q. Shambour and J. Lu, "A trust-semantic fusion-based recommendation approach for e-business applications", Decision Support Systems, vol. 54, no. 1, pp. 768-780, 2012. Available: 10.1016/j.dss.2012.09.005.

[46] A. Felfernig and R. Burke, "Constraint-based recommender systems", Proceedings of the 10th international conference on Electronic commerce - ICEC '08, 2008. Available: 10.1145/1409540.1409544.

[47] O. Medelyan , I. Witten , D. Milne , Topic indexing with wikipedia, in: Proceedings of the Workshop on Wikipedia and Artificial Intelligence: An Evolving Synergy (WikiAI@AAAI), Illinois, UK, 2008, pp. 19–24

[48] S. Kim , O. Medelyan , M.-.Y. Kan , T. Baldwin , SemEval-2010 task 5: automatic keyphrase extraction from scientific articles, in: Proceedings of the 5th International Workshop on Semantic Evaluation (SemEval’10), Los Angeles, USA, 2010, pp. 21–26.

[49] Shu, X. Shen, H. Liu, B. Yi and Z. Zhang, "A content-based recommendation algorithm for learning resources", *Multimedia Systems*, vol. 24, no. 2, pp. 163-173, 2017. Available: 10.1007/s00530-017-0539-8