**Abstraction**

On the subject of the database, all students were given one task, which we need to cope with it by the end of December 1. Our joint task is to compile an application, program, website that will help you determine what specialty to choose before entering the university . To determine the specialty it will take into account your preferences in academic subjects, and will also consider the desired profession . We have assembled a team of three people to carry out this task .

By parsing data from websites we filled necessaries tables. After that we connected our filled tables with sql oracle with which we made our specified program.

Our project is going to help student ,who still don’t know where to go after graduatıng school, by recommending them specialities based on their preferences in academic subjects and on who they see themselves in the future.

**Introduction**

This project is created as a project work on the subject database. The main idea of this work is the creation of an automated program , which will give recommendations for admission to a particular specialty, based on your interests in the subjects and the desired future profession. In this developed program a mandatory component is the database with which we will make input and output of data.

Our program is mainly designed for future students who still can not figure out what specialty and university to apply for. Instead of looking for hours on the Internet on whom to study, we have compiled a program for applicants that will help them with the choice . Using this application will be very easy . You will need to fill in the fields with questions on which the program will make an analysis and display the specialty and the University in which you should enroll. The questions will concern two main subjects in which the future student passed the UNT and the profession in which they want to see themselves in the future.

The data for the program with which will be produced analysis for selection of recommendations were derived from the site through parsing. Information about specialties and respective universities was initially given by the lecturer

was connected with data of professions by codes of specialties.

**Aim and objectives of research**

The main point of our project is creating a program that will help future students with definition their specialty . To do this we collect information by parsing for the table of professions from the site, which has a classifier of specialties and professions.

**Background/Literature Review**

In our project we used high programming language Python. It was chosen because using this programming language is easy , it is convenient to work with data . In addition , Python has special tools , that makes programs work faster . These tools is called libraries(library is the set of functions and methods, that perform many actions,but it doesn’t need to write a code). Examples of libraries of Python , that were used in our work:

* Pandas;
* Python beautifulsoup;

By Python beautifulsoup we were able to parse necessary information of our profession from websites with classified profession and specialities tables .

Pandas(data analysis on Python)makes easy to work with dataset, creates csv files .

The data that we received through parsing was got from special website with classified professions and speciality - <http://edu.gov.kz/m/deyatelnost/detail.php?ELEMENT_ID=664> .

**Methods and Materials**

In this project, we had to get data from classifier. Using Parsing and Programming language, we use Python. To parse, in computer science, is where a string of commands – usually a program – is separated into more easily processed components, which are analyzed for correct syntax and then attached to tags that define each component. We used python’s Beautiful soup library. This library let us get the HTML page with the information and find information we needed. After the parsing data from <http://edu.gov.kz/m/deyatelnost/detail.php?ELEMENT_ID=664> we need to put it to csv file. Using Python’s Pandas library we can put data to csv file. Pandas is Python’s library which let us easily work with data and create csv format file.

We made our Application with java programming language. We use ‘jdbc’ tool to connect oracle database and our application.

Data and Results

In this project we need 3 csv format file:

1. Professions and specialty

1. Profession

2. Specialty

1. Specialties and code
2. Specialty
3. Code of specialty
4. Specialties and university of them
5. Code of specialty
6. First subject
7. Second subject
8. Universities

Our database consists of 3 tables:

1. Professions
2. Specialties
3. Universities

Professions consist of:

1. Profession column
2. Specialty column

Specialties consist of:

1. Specialty’s name
2. Code of specialty

Universities consist of:

1. Code of Specialty
2. First subject needed to enter specialty
3. Second subject needed to enter specialty
4. Universities

First, we created csv format file with columns “specialty” and “professions”, where each profession have the specialty they relate to. For example, we have three profession with the same specialty. It will be written in csv format file in this form:

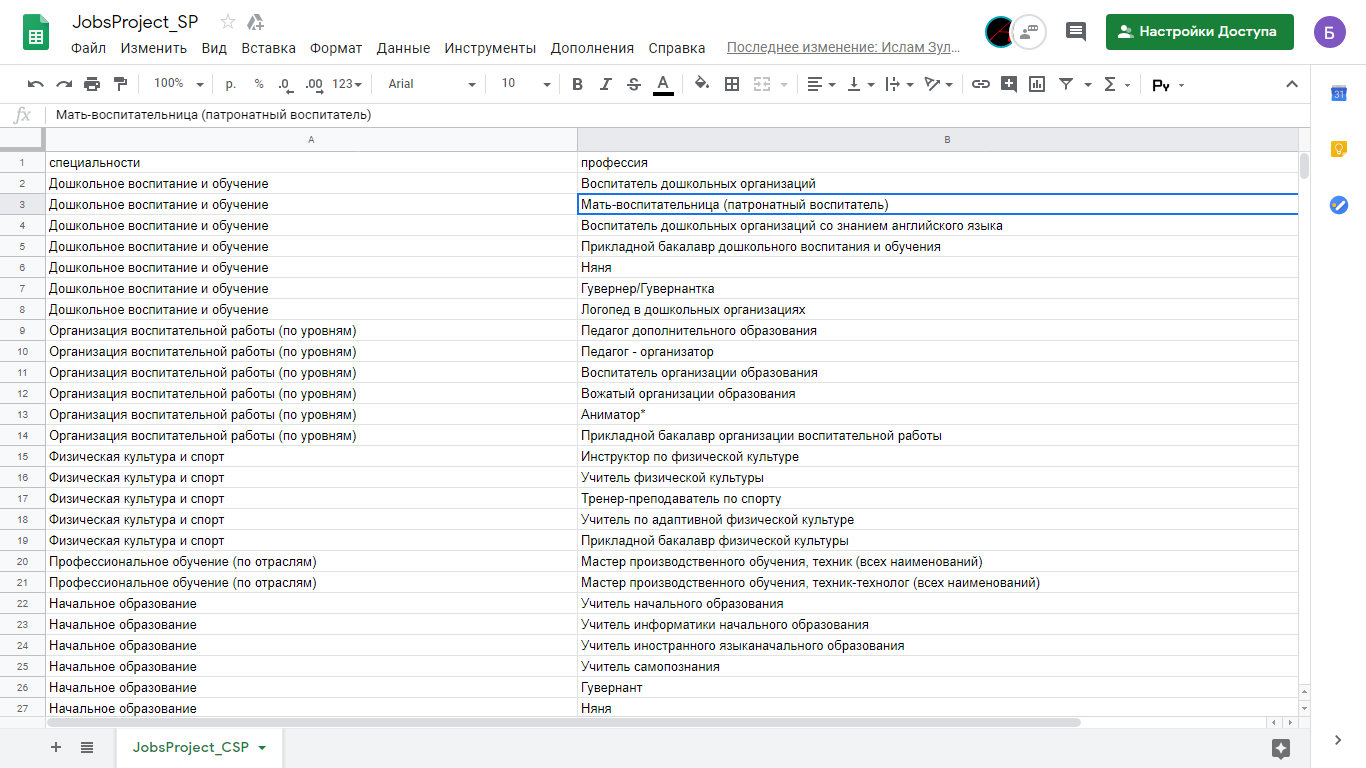
‘specialty1’,‘profession1’

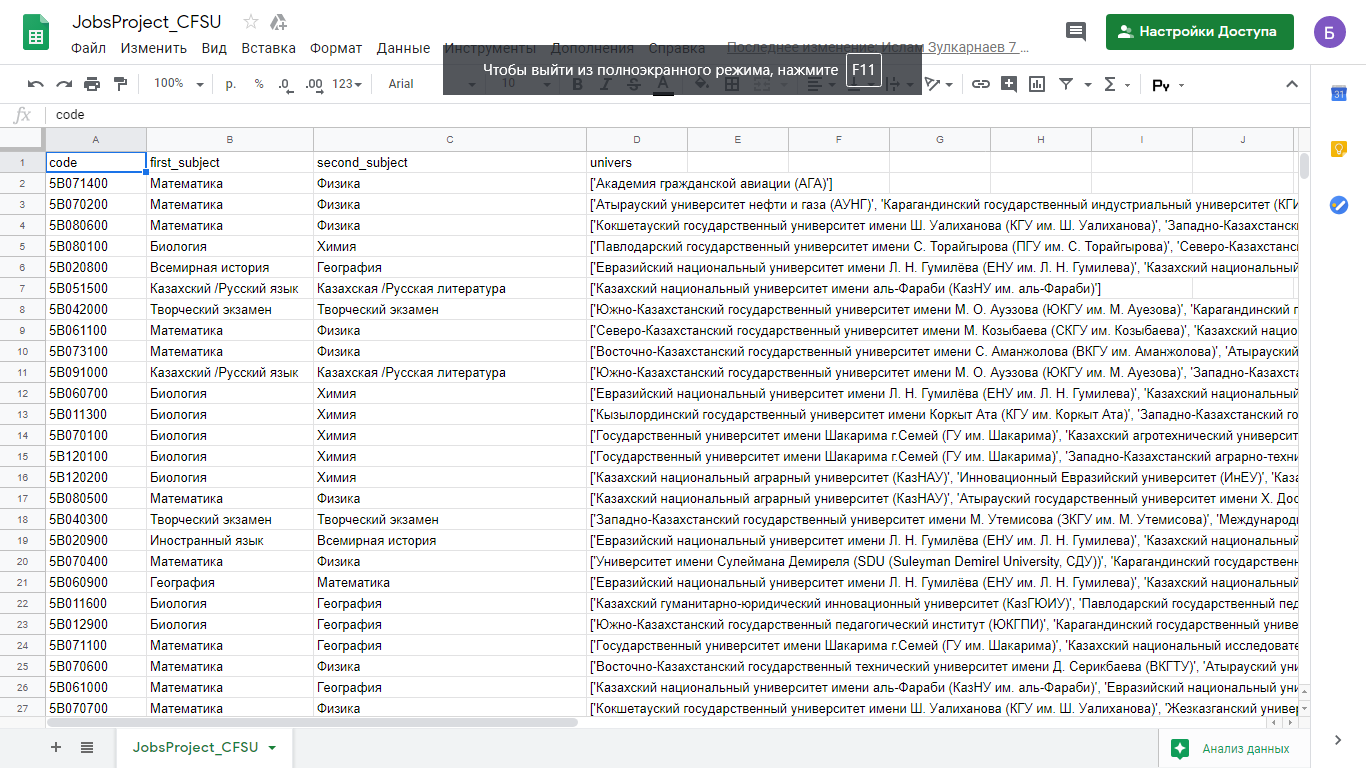
‘specialty1’,’profession2’

‘specialty1’,’profession3’.

We need this information to make recommendation for user. Use choose profession and then we show him specialty he may choose for his study. Specialty is needed for connection between profession table and specialty table in database.

Next, we need to define specialties code. We created new csv file with “Specialty” and “code” with code of specialty in Kazakhstan. This information is needed to connect Universities table and specialty table



The last csv file is universities and their specialty. We parsed this information from [www.univision.kz](http://www.univision.kz) . In this csv file we have to create columns “code”, “first\_subject”, “second\_subject”, “university” , where “codes” means the code of specialty we chose, “first\_subject” and ”second\_subject“ means subjects that student have to pass on national exam, And “university” is the name of universities that have that specialty in their courses. So that we have csv file which looks like shown below: 

We use Oracle Database with PL/SQl. We got 2 PL/SQL procedure:

1. findSpecialty
2. findUniversity

FindSpecialty is the procedure which find all specialties which are connected with profession we need. FindUniversity is the procedure which find all universities that are fit to our request.

This project looks like shown below:

First, user need to select two subjects. For example math and physics. Next user can see list of professions which fit to his two subjects. User must choose one of them and then program will show user list of specialties and universities that he can choose for study. For example user chose programmer, then program will show him specialties CSS and IS, after choosing one of them, user will see list of universities that has this specialties in their courses.

**Conclusion**

Purpose of this project was to help people to choose their profession and university. Project is done and now, we got new skills like work with oracle database with pl/sql. We used in this project our skills like parsing information, work with data and googling. In result improved our skills in database programming. We hope our project will be useful for other people, especially for future students.

**Discussion**

At the begin we had some troubles with the connection between speciality and profession. Then we researched our data collection method. After that all troubles with data was vanished.  As the result, we get clear data which we used in our application.

**References**

Pandas - <https://pandas.pydata.org/pandas-docs/stable/>

BeutifulSoup - <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

Profession classifier - <http://edu.gov.kz/m/deyatelnost/detail.php?ELEMENT_ID=664>