****

**SCHOOL OF IT AND ENGINEERING**

**Senior Design Project**

# PROGRESS REPORT

Project title: ***“..........Morphological Analysis of Azerbaijan Language.............”***

**SDP TEAM:**

Team Member 1: **Fidan Aliyeva / IT**

Team Member 2: **Nargiz Huseynova / CS**

Team Member 3: **Nijat Mursali / CS**

Team Member 4: **Jalal Rasulzada / CE**

Project Advisor: **Samir Rustamov**

SDP Coordinator: **Name / Surname**

## Baku 2018

1. **Objectives:**

In our generation there are several technological advances that help us to create and develop different kinds of machine learning algorithms in different areas in order to make our life easier. For instance, we can find lots of algorithms that are already made by developers in order to perform the morphological and lexical analysis in different languages, however, such an algorithm has not been created for our national language. The problematic issue is the lack of the morphological analysis of Azerbaijani linguistics currently and there is not such an avaiable source which can include the morphological analysis of every word in Azerbaijani language in order to eliminate that problem. There are lots of advantages of our algorithm in order to solve many problems, therefore our team has decided to make the vision on this topic in order to perform the tasks for Azerbaijani language Thus, since the morphological analysis of Azerbaijani language has not been completed by any other source including Microsoft Word with its editing function or other applications due to not considering every single word according to its right spelling, our basic objective is to radically solve this issue.

1. **Significance:**

Thanks to advanced software for word correction in modern times writing in almost all different languages is made very easy especially for those who are writing long academic papers and articles. However, as mentioned before this is almost impossible in our national language and the available software is far behind the perfect. Unfortunately, in this modern world, people writing papers in Azerbaijani language have to use old-fashioned ways such as books and dictionaries to improve their writings. Our project aims to solve this problem and take Azerbaijani language to the same level as others. One of the outstanding potential applications is word correction which will do its job very well by our entered algorithms and satisfy the potential users of the application. The target segmentation can be demonstrated as the essay and article writers, journalists or anyone else that are involved such kind of academic issues whom our website will definitely help.

1. **Related Work/Novelty:**

In today’s world there are several tools and algorithms that were made by individuals to develop and enhance the language and make our life easy. One of such algorithms is to demonstrate the possible options for any word and check the words in the sentence, however, such an algorithm has not been developed for the Azerbaijani language. There’s one tool for checking the words only, however, that algorithm is not able to check it for the sentences. There are several problems that can be solved by implementing this algorithm. First of all, we can and will be able to contribute to the linguistics by implementing this algorithm by showing possible ways to generate different words for any word. Therefore, our team has decided to implement this algorithm within Azerbaijani language. Here are the main approaches that we will try to practice: checking the possible cases for every word, checking the sentence and dividing it into parts for every word. Here are the main project objectives:

1. Getting and adding the databases of information related to Azerbaijani language
2. Checking the database of words every time
3. Applying our algorithm for every word such as noun, verb, adjective and pronoun
4. Evaluating our algorithm with other algorithms by checking the similarities and differences
5. Designing the website for our algorithm in order to make other users evaluate our algorithm

Our team has already checked the possible algorithms that made in English and Azerbaijani language. Our job will be applying these approaches into Azerbaijani language and get help from linguistics to improve it. Till this time, our team has already made research related to this issue and studied several algorithms in order to help us accomplish this project.

The significance of our project will be that we will contribute to the language and linguistics by applying this algorithm and there’s no powerful tool to check the words and sentences. We already discussed these issues with Dr. Samir Rustamov who is our project advisor. We will be the ones who develop such kind of tool in Azerbaijan by the help and advice of our project advisor Dr. Samir Rustamov and SDP coordinator Dr. Abzatdin Adamov.

1. **Chosen Solution:**

For more than a century, many investigations have been made by researchers concerning about linguistic algorithms for different languages, especially for English. For making this kind of algorithm related to Azerbaijani language, we had to make a deep investigation for the algorithms that have been made for other languages and recognize how they work. To understand their strength and weakness would help us to develop better algorithm suitable for Azerbaijani language. As mentioned above, we have already made several researches in order to understand other algorithms and work principle of them would make us make it for our language. As we all know, Azerbaijani language is one of the richest languages in the world because of the number of words and different variations of those words make it even better. Another problem for implementing this algorithm was that in our database we have more than 40.000 Azerbaijani words that were taken from websites, books and articles, however, this number is not enough for getting the exact result for the algorithm. We have also found the solution for this issue which was to get more books from the linguistics of Azerbaijan and add them to our database. Another problem was that in Python programming language we don’t have “ə” character and our solution was to copy it from some text and add it to the algorithm. Another problem was changing of one character if we add prefix to the word. For example, let’s take “get” verb and if we add “iş” which is the suffix that creates noun from verb. The problem here was that “t” character should change to “d”. Fortunately, we created the script for that and solved the problem.

For our algorithm in Python environment, we added all possible suffixes for the noun, verb and adjective till this moment and we will do our best to finish all of the suffixes for every word like adverb and pronouns.

1. **Output/Deliverables:**

In most of the languages there are such tools that perform morphological and lexical analysis by using the words and even sentences. However, in our main language which is Azerbaijani we don’t have tools for checking the words. In our thesis, we will search about the methods that were using for other languages such as English language and apply those analyses for Azerbaijani language. For instance, imagine the user adds any word and what the program will do is it will search the word and divide it into parts so-called tokens.

Since the website will be ready as an outcome, the users can include the word, sentences, or composition, and the website will automatically make the morphological analysis of entered input.

1. **Project Management/Timeline:**

Initially, the Azerbaijani words were classified and dictionaries were improved. Gradually, noun and verb compositions were developed and prepared to be included. Then the algorithms for further implementations are ready. Firstly, we developed algorithms and gathered significant data about morphological and lexical structure of Azerbaijani language. During our researches, we observed that noun and verb have the most complex and challenging structure, therefore we decided to start with them. We gathered and prepared data about the noun and verb and made them ready for implementation. Also, we improved the dictionary and broke it down into categories to boost the performance of our software. After implementation of noun and verb we tested and debugged the written code.

Right now we are preparing and gathering information about other parts of speech – adjective, pronoun, numeral and adverb for implementing and developing a new interface. On January and February we are going to devote our time for implementation, testing and debugging them. After that we are planning to develop a website for our software and make it available for public usage. Also, we are aiming to get help from scholars from higher institutions such as Azerbaijan University of Languages to make our work more precised and accurate.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Research |  |  |  |  |  |  |  |  |
| Implementation |  |  |  |  |  |  |  |  |
| Debugging |  |  |  |  |  |  |  |  |
| Implementation |  |  |  |  |  |  |  |  |
| Testing |  |  |  |  |  |  |  |  |
| Website |  |  |  |  |  |  |  |  |
|  | October | November | December | January | February | March | April | May |

1. **What Done?:**

The filtering of Azerbaijani language, morphological analysis of the noun and verbs and etc. have been achieved up to now. Each member of the team has contributed lots of time and effort to successfully establish the framework and reach our goals with their assigned tasks.

Fidan Aliyeva and Nargiz Huseynova had been gathering important data about the morphological structure of Azerbaijani language that we needed during our implementation. Moreover, they had been performing the testing phase of the written code in order to detect and report bugs. Apart that Fidan Aliyeva is working on development of website for our project. The main duties for Nijat Mursali and Jalal Rasulzada were organizations and research of the algorithms and implementing them.

All the words in the dictionary have been broken down into categories and groups to increase the efficiency of the program. Our program can already analyze nouns and verbs and in any form and demonstrate accurate morphological analyses of them. Also, by entering a given noun or verb, it prints out all the possible forms of word with prefixes and suffixes. Most of the bugs and lexical mistakes occurring during the usage are investigated and fixed. The algorithms for further stages of implementations are ready.

Students should not write in the section below:

**Project Advisor:**

Project Advisor’s comments on Progress Report:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Recommendation:

Passed Failed

Project Advisor: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Signature & Date \_\_\_\_\_\_\_\_\_\_\_\_ …./.…/2018

**SDP Coordinator:**

Approved \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ SDP Coordinator …./.…/2018

Conditional Approval: Recommended Revision:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Disapproved: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Reason for Disapproval: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_