International Journal of Soft Computing 10 (5): 348-352, 2015
ISSN: 1816-9503

Design of Virtual Keyboard for Tatar-Speaking Users on the Basis of the Mobile Operating System Android

Andrey Vladimirovich Danilov and Lyaylya Leonardovna Salekhova
Kazan Federal University, Kremlevskaya Str. 18, 420008 Kazan, Russian Federation

Abstract: The study presents the design of the virtual keyboard for the Tatar language on the basis of the mobile operating system Android. The researchers have analyzed the software products of the similar subject matter, specified the positive and negative features of the keyboards as well as studied the technologies allowing simplifying and speeding up the process of the keyboard entry. The result is that the main directions of the software product design have been determined: optimization of the style and symbol arrangement as well as design of the predictive entry system. The study describes the keyboard components designed and their functions. Submited design keyboard informed and shows location letters of keyboard. The keyboard style is presented; the process of design of the predictive entry system is described. The presented predictive entry system knows no equals among the virtual keyboards for the Tatar language. The keyboard designed was placed in the Internet-store of mobile applications Google Play Market and is distributed as free software. The keyboard has received positive feedback; the average grade of the application is 4, 3 out of 5. This product is marketed as an example of an information solution aimed at preservation and development of the Tatar language in the social-humanitarian area.

Key words: The Tatar language, localization, Android, keyboard, predictive dictionary

INTRODUCTION

The Republic of Tatarstan is a multi-national country populated by representatives of different cultures. According to the All-Russian population census in 2010 the Tatars made 53.15% of the population of the Tatarstan Republic (Anonymous, 2010). For preservation and development of the Tatar language it has to become the language of communication in the info communication environment (Zaripova, 2014) which requires localization of the software product and application interfaces. This task is being solved step by step (Nevzorova *et al*., 2013; Anonymous, 2015a, b), however, it still remains acute for mobile technologies. Absence of the quality software resulted in the use of the Russian keyboard layout by most of the Tatar-speaking users. This study presents the process of design and popularization of the virtual keyboard as an example of an information solution aimed at preservation and development of the Tatar language in the social environment.

MATERIALS AND METHODS

In order to design the virtual keyboard for the Tatar language the following tasks shall be solved:

- To analyze the market for applications of the similar functionality; to generalize the data obtained and to specify the positive and negative aspects of the existing software products
- To design a new software product using the existing helpful experience
- To place the finished software product in the mobile apps store Google Play Market

In order to solve the first task the virtual keyboards for the Tatar language available on Internet have been found and analyzed. In general, the software scope is similar since the same principle of matching the letters and symbols with certain keys is used. The analysis of the keyboard ergonomic features was performed according to several criteria in particular:

- Arrangement of the Tatar symbols at the keyboard
- Availability of the predictive dictionary

Let's review the results of analysis received on the basis of the first criterion. The symbol layout determines the symbol typing rate and ease of use of one or another key. The arrangement of symbols in different software products is presented in Fig. 1-4. In the virtual keyboards analyzed the two main approaches to the arrangement of the Tatar symbols have been used:

- The first one; allocation of symbols in a separate line (Fig. 1 and 2)

Corresponding Author: Andrey Vladimirovich Danilov, Kazan Federal University, Kremlevskaya Str. 18, 420008 Kazan, Russian Federation

Fig. 1: The JBak keyboard for the Tatar language

Fig. 2: TatarKey keyboard

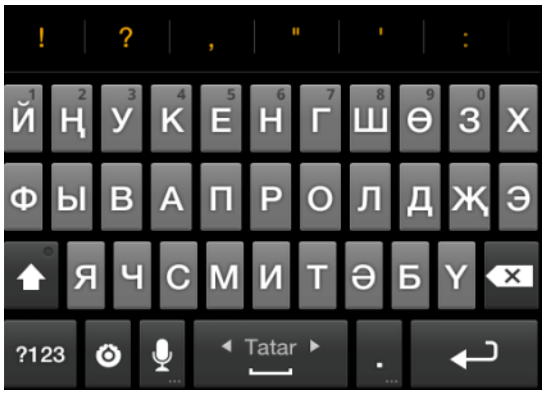

Fig. 3: GingerBread keyboard for the Tatar language

- The second one; arrangement of symbols in a way that is similar to the QWERTY (**ЙЦУКЕН**) layout in the Tatar language (Fig. 3 and 4)

Both feature limitations: thus, the use of the first approach results in the lack of space on the mobile device display which makes the entry process inconvenient. By using the second method one may arrive at the conclusion that the keyboards are not convenient enough

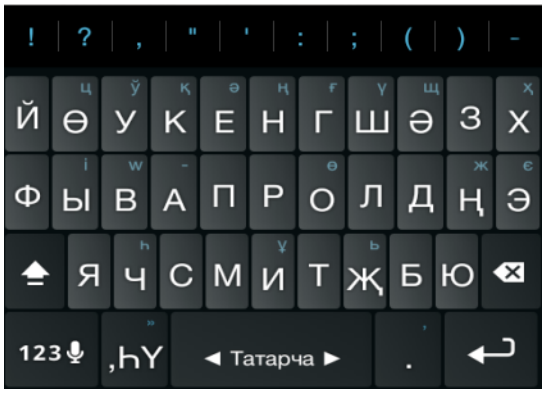

Fig. 4: Multilang keyboard for the Tatar language

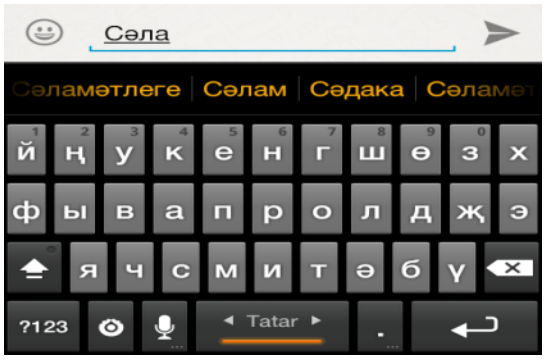

Fig. 5: Functioning of the predictive system on the GingerBread keyboard for the Tatar language

as there are 6 letters more in the Tartar alphabet than in the Russian one this is why one key shall refer to 2 letters for example in **ЙЦУКЕН**-layout the letter "**Н**" is placed above the letter "**Ц**".

Let's review the results of analysis obtained on the basis of the second criterion. The predictive dictionary has a significant effect on the typing rate allowing the user to entry words from a special panel without typing them fully. The principle of operation of a keyboard with the use of a predictive dictionary comes down to the fact that the program analyzes the letters entered by a user and suggests appropriate words from the predictive dictionary which accelerates the word entry significantly. The scope and principle of the dictionary arrangement predetermine its quality. In the software programs, being examined such a dictionary was composed for one keyboard only. The predictive dictionary consisted of 60,000 words and their frequency characteristics did not correspond to the lexicon that is commonly used by mobile communication. Thus, the frequently used word "**сәлам**" appears on the panel only by the keyboard entry of the fourth symbol, i.e., when "**сәла**" is entered (Fig. 5). In the researchers opinion by

optimizing the predictive dictionary, orienting it towards the mobile slang one may achieve substantial improvement thereof.

RESULTS

The result is that the following directions of the software product development of have been defined. Firstly, the task of optimization of keyboard layout was set. Secondly, the necessity of creation of a predictive Tatar dictionary oriented to the mobile slang. The author's concept consisted in dividing the predictive entry system into two subsystems. The first subsystem that was named during the development process the predictive entry system one (SPV-1) functions similarly to the predictive dictionary, i.e. by entry of symbols the system suggests quick entry of the most frequently used words. The second subsystem (SPV-2) analyzed the entered word, then suggests the established Tatar expressions (cliches) included in the predictive dictionary.

The second task included development of the application and predictive dictionary for which a team of development from the research fellow of the faculty of mathematical linguistics and information systems in philology at the Kazan Federal University and the Institute for Applied Semiotics under the Academy of Sciences of the RT was created. The team was divided into groups according to the work specifics. The first group worked at the program code and the software product design, the second one composed the predictive dictionary.

Design of the virtual keyboard for the Tatar language on the basis of the mobile operating system Android included the following stages:

- Design of the style and defining the key layout
- Construction of the algorithm of operation of the keyboard operation and predictive entry system
- Composing the dictionaries according to the algorithm being designed
- Recording the code, inclusion of dictionaries and compiling the application
- Testing the application

On the basis of the first and the second stages, the layout of the Tatar symbols on the virtual keyboard that is presented in Fig. 6 was selected and the basic principles of the keyboard operation were developed.

As the result of the adding the Tatar letters and lack of space on the display some characters have been hidden (Table 1).

This principle of the character substitution arose as a tradeoff between sparing the display space and frequency of the characters use: the less frequently used

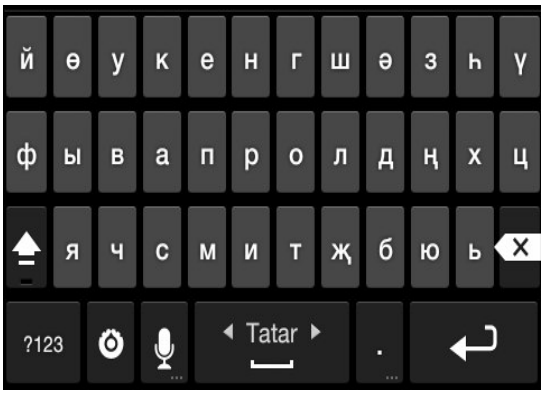

Fig. 6: Final version of the Tatar keyboard layout

Table 1: The Tatar letters and hack of space on the display some charactors

Symbol on the keyboard	Hidden symbol
Ә	Э
Ш	Щ
Ь	Ъ
Е	Ё
Ж	Ж

Table 2: The scheme of the database for the predictive entry subsystem SPV-1

Word	Word frequency
сәлам	19000

Table 3: The scheme of the database for the predictive entry subsystem SPV-2

Word_1	Word_2	Word frequency
ничек	хәлләр	13000

Moreover, the new keyboard included the Russian **ЙЦУКЕН**-layout and the English QWERTY-layout. This was determined by the features of operation of the mobile operating system Android. OS recognizes keyboard as a standalone application that is started or activated automatically. Thus if the keyboard designed had only the Tatar layout the user would need to enter the settings menu to select another keyboard supporting the Russian or English layout. Since, these layouts are used frequently the decision was made to include them to the product being developed.

Following the third stage of the work, the algorithm of operation of the predictive system was developed. The system is divided into two subsystems SPV-1 and SPV-2. Each subsystem refers to the own tabular database where the words and expressions as well as frequency of use are indicated. The schemes of Table 2 and 3 are presented. Principle of operation of the predictive system is presented in Fig. 7.

At the fourth stage, the necessary dictionaries for the predictive system were composed. The search and

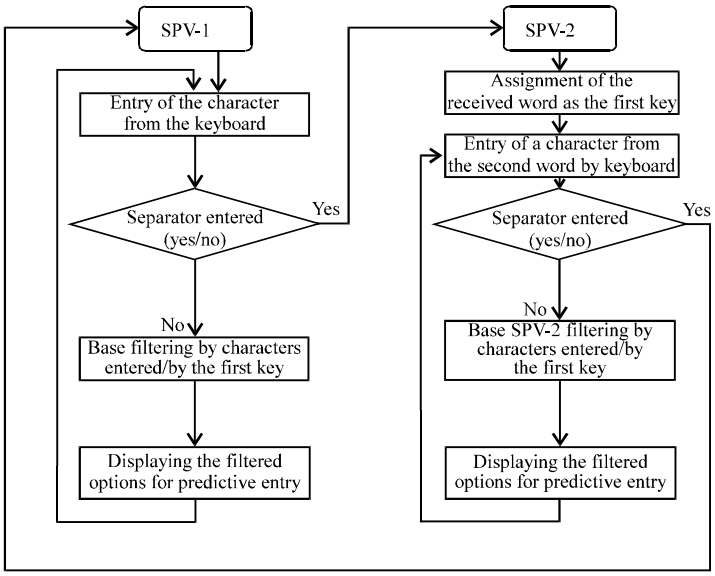

Fig. 7: Algorithm of operation of the predictive entry subsystem

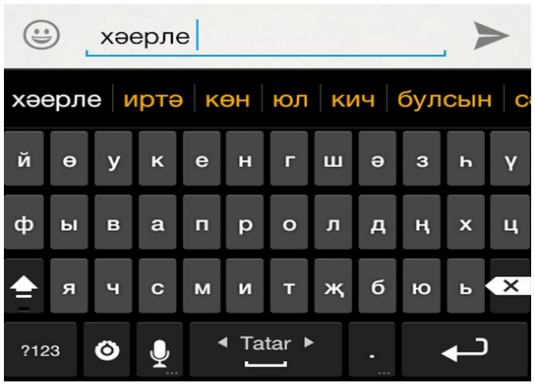

Fig. 8: Functioning of the application and the predictive entry system SPV-1. By entry of the characters "**хәе**" the subsystem suggests the popular variants for speed typing

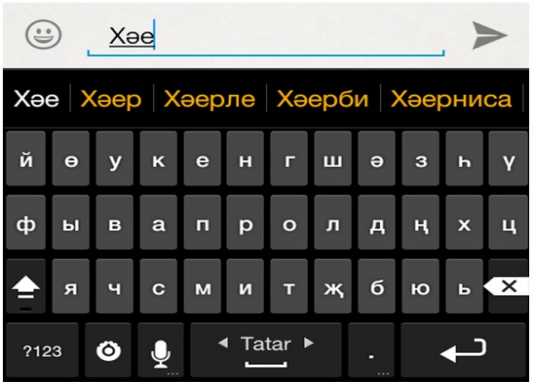

Fig. 9: Functioning of the predictive entry subsystem SPV-2. By entering the word **хәерлe** the system suggests a few alternate expressions already

analysis were performed with the use of Internet-resources. As the result of this, the two dictionaries containing about one million of words and expressions have been obtained.

At the fifth stage, the draft versions of the dictionaries were included in the software product after which the code was developed and the predictive system was tested. It was necessary to optimize the code and the dictionaries relying on the three criteria: the application capacity, speed and dictionary volume. As the result of optimization the algorithm of the predictive system operation was improved and the dictionaries were reduced by 50%.

Reduction of the dictionary volume does not affect the dictionary performance as the least frequently used words and expressions were deleted. However, the amount of such elements constitutes the greater part of the base; the deletion reduced the capacity of the memory occupied which gave a benefit to the developers. Also at this stage, the first test version of the application was compiled. The process of working with the application is presented in Fig. 8 and 9.

At the sixth stage, the application was tested thoroughly, the errors were detected that have been corrected thereafter. The third task consisted in posting the application at the site for distribution of the software

products for the operating system Android Google Play Market. After release this software product became available for all users of the OS Android.

CONCLUSION

As of the date of the paper publishing the virtual keyboard for the Tatar language on the basis of the mobile operating system Android had been designed. Release at the Google Play Market site took place on December 8, 2014. The application is called "**Тиз.Яз**" and is freely accessible through Internet. During the period from December 8, 2014 till March 10, 2015 the application was downloaded by 2124 users.

There has been developed the software product promoting to the enhancement of the area of use of the Tatar language as the language of info communication technologies.

ACKNOWLEDGEMENT

The study was prepared within the frameworks of implementation of the "Plan of measures on implementation of the Program of improving the competitiveness of the FSAEI HVE "K(P)FU" among the leading international research-educational centers for the years 2013-2020.

REFERENCES

Anonymous, 2015a. Localized software products. Tatsoft. http://tatsoft.tatar/ru/prilozheniya.

Anonymous, 2015b. Translation of Android-programs into the Tatar language. Mobile Applications Portal 4PDA. http://4pda.ru/forum/index.php?showtopic=475365.

Anonymous, 2010. Informational materials on the final results of the All-Russian population census in 2010. Web-site of the Federal State Statistics Service. http://www.gks.ru/free_doc/new_site/perepis2010/perepis_itogi1612.htm.

Nevzorova, O., D. Suleymanov, R. Gilmullin, A. Gatiatullin and B. Khakimov, 2013. Tatar National Corpus "Tugan tel": structure and features of grammatical mark-up. Corpus Resources for Descriptive and Applied Studies. Current Challenges and Future Directions: Selected Papers from the 5th International Conference on Corpus Linguistics (CILC2013) Ed. Chelo Vargas-Sierra. Procedia Soc. Behav. Sci., 95: 68-74.

Zaripova, R.R., 2014. Definition of development level of communicative features of mathematical speech of bilingual students. Salekhova, L.L., N.K. Tuktamyshov and R.F. Salakhov (Eds.). Life Science Journal, No. 11, Vol. 8. http://www.lifesciencesite.com/lsj/life1108.