Research Article

SOFTWARE AND DATA ENTRY PRINCIPLES OF THE ONLINE DIALECT DATABASE

Kholova Muyasar Abdulhakimovna
Doc.Of Phil.Scien.(Phd) Termez State University, Senior Lecturer Of The Department Of Uzbek Linguistics Termez City Of Surkhandarya Region, Uzbekistan

ABSTRACT

In this article we will talk about the formation of a data warehouse for an online dialect database, about the processes of sequencing the import of data warehouse tokens for an online dialect database. In addition, a data warehouse table was prepared, as well as common features and differences in the presentation of the dialect database, as well as the software structure and software of the online dialect database.

KEYWORDS

dialect database, dialect expedition, file, Microsoft Exsel program, software, dialectological package, dialect location in the explanatory dictionary of the Uzbek language, transliteration, transcription, dialect annotation, dialect example, category, statistics, own and mastered layer, insertion into the database.

INTRODUCTION

Analysis and results It is known from the experience of dialectal corpora of the world that it is impossible to start corpus practice without forming a base of dialects. The formation of the base includes several stages. The first stage requires manual work, that is, it is necessary to form a data archive. Dialectal expeditions will form a database in the regions. Dialectal database is a collection of lexical units of the language of

several small regional ethnic populations using the same literary language, classified in artificial intelligence, which presents the possibilities of online use and sorting. The difference of this database from literary language corpus databases is in the transcription presentation of dialect units.

LITERATURE ANALYSIS

The International Dialects Archive of English[11] is one of the most prominent databases for collecting dialects in one database or corpus in the world, while the Bashkir dialect database offers: lexicological database, dialectological map database, text database [8.10]. E. V. Paducheva should express an opinion about the collected linguistic data for research, that most of the examples of any research are given on the basis of dictionaries, if it is not possible to cover the collected data on the basis of examples, then from text examples indicates that it should be used[3.

5]. When it was not possible to use dictionaries in the explanations and examples of the facts of the dialect, an attempt was made to explain them with examples from the everyday life of the ethnic people.

RESEARCH METHODOLOGY

Importing data into the dialect database was carried out using the methods of "observation", "conversation", "study and analysis of documents", "sequential placement".

ANALYSIS AND RESULTS

i.e. conversations of dialect representatives were observed and questionnaires were filled out based on the conversation. Dialect lexemes presented in works of art are collected, studied and placed in the table of Excel program of Microsoft Office. The sequence of data import into the online language database is as follows:

1.The provided sample file is downloaded.

Picture 1.

1	2	3	4	5	6	7	8	9	10	11	12	13
Adabiy til	Transliteratsiya	Transkripsiya	Turkum	Sheva izohi	Misol	Sinonim	Omonim	Antonim	O'zlashgan qatlam	O'zlashgan qatlam Izohi	Rus tili	Ingliz tili

A table sample prepared for uploading to the dialect database.

1. In the file, the corresponding cell is filled with a lexeme related to the dialect [2. 45], when the number of rows is not enough, the next database should be entered through another copy.

2. If the use of lexemes used in OTIL is observed in the database of dialects, the database does not automatically create the need to enter using it [1.225], that is, there is no need to enter the word every time. If there is a new lexeme, the program loads it into the database. If the participation of this word in OTIL is not observed, the cell is left open. It is envisaged that the lacunae will be sorted later.

Picture   2

1.   An example of a step 3 process for importing data into an online language database.

2. The transliterative form of the dialect is used if it is a previously introduced word, if it is a new word, the dialect is added to the table of words

3. At the next stage, a base of words related to the dialect is formed, with the transscriptive unicodes of the dialects correctly defined [5.20]. In this case, when the base matches the previous data, it works on the basis of the commonality of the words in the cells, and the sorting stage is applied to this case.

4. If there is an explanation of the word in OTIL, it is entered first in the cell of dialect explanation. Unexplained lexemes are given their own meaning based on their content. There are many lexeme explanations used in the figurative sense. is represented in the cell by the abbreviation

5. An example is included in order to express the content of the dialect in the context and to reveal the substantive and figurative meanings, since the lexemes in the base are intended to cover not only the spoken language, but also all the dialects that exist in the spoken and written language, the interpretation of the examples in the works of art , therefore, it is planned to include lexemes characteristic of the oral communication style[7].

6. At this stage, it is better to prepare and load the base of all the cell types of dialects once, otherwise the database itself will be able to choose the type for the given example. The sample file is listed and sortable. If the user wants to research one category available in the database separately from others, he can do it by using the button for selection and sorting in the database.

Picture 3

An example of the 8-step process for importing data into an online language database.

2. Synonym entry process: synonyms should be written in the synonym column, separated by a ',' sign: synonym1, synonym2, and the explanation should be written in the next column: A.t. Then Synonym 1 (A.t.), Synonym 2 (A.t.) will be displayed after import[8].

3. Synonym entry process: synonyms should be written in the synonym column, separated by a ',' sign: synonym1, synonym2, and the explanation should be written in the next column:

Picture 4

A sample of the 10-step process for importing data into an online language database.

Then, after import, Synonym 1 (A.t.), Synonym 2 (A.t.) will be displayed. The base presentation presents the two cells as a single cell[4].

2. Homonym entry process: similar lexemes should be written in the homonym column, separated by the sign ',': homonym I, homonym II are entered and presented in this form.

3. The process of introducing antonyms: lexemes similar to the antonym column should be written separated by the sign ',': the sequence of antonyms is represented by ".".

4. For the cell intended to distinguish whether the lexemes belong to the own or the acquired layer, no entries are intended for the units belonging to the own layer. In the input process, two cells are designed for the mastered layer, and the lexeme layer is entered in the first cell in the form of keywords such as Persian, Arabic, Persian+Uzbek, Mongolian. Separation of cells is provided for sorting the program and is carried out using the same keyword. The second cell contains the explanation of these lexemes and presents the two cells as a single cell.

Figure 5

	H	I	J	K	L	L	M
						O'zlashgan qatlam Izohi	Rus tili
1	Sinonim Izoh	Omonim	Antonim	O'zlashgan qatlam			
2	Столбец2	Столбец3	-	Столбец4		Столбец5	Столбец1 3
3 4	Izoh	omonim	antonim	O'zlashgan qatlam		Izohi	Rus tili
5			-				шелуха
6		I. savdo, g'avg'o.		fors		[sho'r bosganlik, tuzi ko'plik]	
7 8	tuzli		Shirin, tolli	fors		[sho'r bosganlik, tuzi ko'plik]	солененьк ий
							Справляе

13 is an example of a sequence for importing data into an online language database.

2. Russian and English translations of Sheva should be separated by a `,' sign. In this case, separate cells of the words are separated and entered.

Figure 6

Example of process 14 for importing data into an online language database.

2. The internal software of the language database is implemented in the universal character encoding standard "UTF-8". All information is provided on the basis of the Boisun language database prepared by the author.

Figure 7

15 is an example of a sequence for importing data into an online language database.

Software of the online dialect database on a computer is implemented as follows:
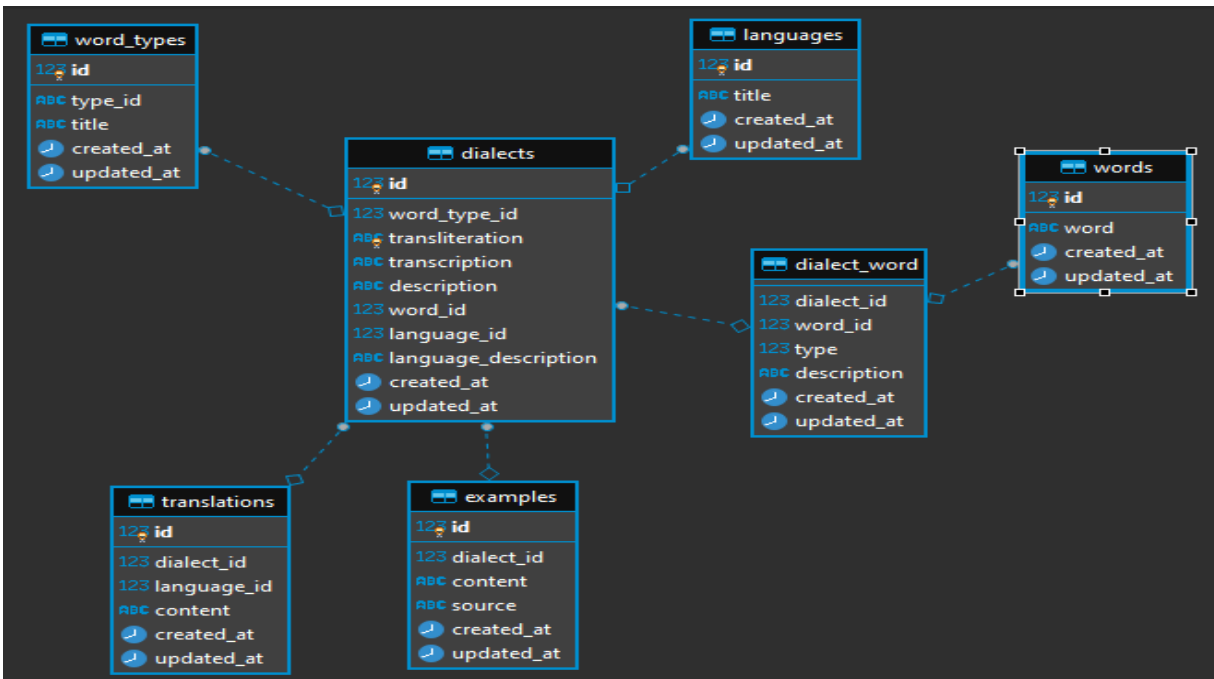
Word types:

CREATE TABLE `word_types` (

`id` bigint(20) unsigned NOT NULL AUTO_INCREMENT,

`type_id` varchar(50) COLLATE utf8mb4_unicode_ci NOT NULL COMMENT '1: Independent, 2: Auxiliary, 3: Separate',

`title` varchar(50) COLLATE utf8mb4_unicode_ci NOT NULL,

Figure 8

Note: Internal integration of posting new dialects in the online dialect database

After completing the above 15-part process, it is tabulated and prepared for loading into the dialect database.

Figure 9

Note: before importing into the online dialectal database

The bottom line is that the units of the existing collected dialects are uploaded to the online dialect database at once, partly with the help of a human factor. By implementing this process, dialectology and corpus linguistics researchers will be able to:

use of dialectal units in three different ways at the same time;

conducting research based on any selection from the database;

the existence of synonymous, homonymous and antonymous relationships of literary language and dialect lexemes is a source for research not only in linguistics, but also at the level of lexicology of linguistics;

the possibility of selection in the studies carried out on the genesis of dialect lexemes;

analysis on the basis of generality and specificity in the comparison of dialects of the common Turkish language;

translation of dialect lexemes into Russian and English;

at the same time, the problems of creating new generation dictionaries are urgent in linguistics, within the framework of lexicology, without giving an explanation of the words, but assessing the possibility of comparing and comparing them[8] at the same time these dialects are a dictionary[11], we can say that it can fully meet the requirements of such an electronic dictionary.

REFERENCES

1. Abdulhakimovna X. M. Baysun dialekt corpus: Genesis (Baysun district is in the case of "J" dialect) //Asian Journal of Multidimensional Research (AJMR). – 2020. – Т. 9. – №. 2. – С. 224-228.

2. Пўлатов А. Компьютер лингвистикаси. – Тошкент: Академнашр, 2011.

3. Падучева Е. В. Высказывание и его соотнесенность с дей- ствительностью (референциальные аспекты семантики ме- стоимений).М.: Едиториал УРСС, 2004. - 228 с.–С.

4. Mengliyev B., Hamroyeva Sh., Ahmedova D. Kompyuter leksikografiyasi. – Toshkent, 2021.

5. Захаров В.П. Корпусная лингвистика: Учебно-метод. пособие. –СПб., 2005.

6. Ҳамроева Ш. Ўзбек тили муаллифлик корпусининг лингвистик асослари. Монография. Нашр: GlobeEdit, 2020.

7. Холова М. А. O'zbek shevalarining standart tilga munosabati va til tizimidagi o'rni (Boysun tumani "j" lovchi shevalar misolida) // Международный журнал искусство слова. – 2020. – т. 3. – №. 5.

8. Xolova M. ўзбек миллий шевалари корпусини тузишнинг лингвистик асослари (Бойсун тумани "ж"ловчи шевалари мисолида): Филол. фан. фал. док. (PhD) дисс. – Термиз, 2022.

9. Холова М. А. Boysun shevasi korpusi: Ibtido (Boysun tumani "j" lovchi shevalar misolida) // Инновации в педагогике и психологии. – 2020. – №. SI-2№ 5.

10. https://www.dialectsarchive.com/

11. https://dialect.uz/