Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 06, June-2023

The solution to sort entries in the Vietnamese-Ede bilingual vocabulary database

Le Hoang Thi My
University of Technology and Education
The University of Danang
Danang, Vietnam

***Abstract*— The process of querying data in the vocabulary database, the work of arranging the data in ascending or descending order of each language is a criterion that should be considered in the study of building a vocabulary database. The implementation of sorting an English data table in alphabetical order with the Order by clause in the SQL statement is simple. Because the ASCII encoding and database management systems are used with the English alphabet. However, with the ethnic minority languages of Vietnam in general and the Ede language in particular, the implementation of alphabetical ordering has not been supported and has not received the attention of scientists. Therefore, when writing a database programmer for Ede language, it is difficult to present a data table arranged in alphabetical order. To solve this problem for Ede language in the lexical database, the article proposes a solution to sort Ede entries in the Vietnamese-Ede bilingual lexical database, in order to contribute to the search and control. investigate and manage data, build Ede data tables in alphabetical order of Ede language..**

Keywords: Ede language processing, Unicode encoding, entry sorting, vocabulary database, data query

I. INTRODUCTION

All information processing activities on computers are related to text editor. Sorting is the process of rearranging the elements of a certain set of objects in a certain order such as ascending or descending for a sequence of numbers, alphabetically for words. Sorting work is often applied in Informatics applications with purposes such as: sorting data in computers for convenient searching, arranging processing results to print out on reports . To solve this problem for Vietnamese and Ede languages in the Vietnamese-Ede vocabulary database [3], [4], The paper proposes a solution to arrange items in the Vietnamese-Ede vocabulary database, the content of the solution is as follows:

- First, encode Vietnamese and Ede letters into a continuum to allow string comparison in programming applications.
- Moves the encoded entries into the array. Sort the array in alphabetical order.
- Pass the index of the array after being sorted into the sorted index property in the datastore corresponding to the item decoded from the array.

Thus, later when performing sorting of items in the data warehouse instead of sorting by item attributes, we perform sorting by the sort index attribute.

II. METHOD OF ENCODING VIETNAMESE AND EDE LETTERS

A. *Encoding Vietnamese letters*

Each Vietnamese letter will be mapped into a continuous region in the Unicode en-coding. Areas selected for mapping range from 1F00:1F5E. The basis for choosing this region is because this is a continuous area containing characters and characters in this area do not appear in Vietnamese documents. Table I is a table that maps Viet-namese letters in alphabetical order to the extended Greek character area in the Unicode encoding.
For example, the entry from the school is encrypted ῞Hἤά □ὅ῍Ο῍Ι῞Η.

TABLE I. MAPPING VIETNAMESE LETTERS INTO THE EXTENDED GREEK CHARACTER AREA

Vietnamese characters								
a	à	ả	ã	á	ạ	ă	ằ	ắ
ẵ	ắ	ặ	â	ầ	ẩ	ẫ	ấ	ậ
b	c	d	đ	e	è	ẻ	ẽ	é
ẹ	ê	ề	ể	ễ	ế	ệ	f	g
h	i	ì	ỉ	ĩ	í	ị	j	k
l	m	n	o	ò	ỏ	õ	ó	ọ
ô	ồ	ổ	ỗ	ố	ộ	ơ	ờ	ở
ỡ	ớ	ợ	p	q	r	s	t	u
ù	ủ	ũ	ú	ụ	ư	ừ	ử	ữ
ứ	ự	v	w	x	y	ỳ	ỷ	ỹ
ý	ỵ	z						
Extended greek character area								
ἀ 1F00	ἁ 1F01	ἂ 1F02	ἃ 1F03	ἄ 1F04	ἅ 1F05	ἆ 1F06	ἇ 1F07	Ἀ 1F08
Ἁ 1F09	Ἂ 1F0A	Ἃ 1F0B	Ἄ 1F0C	Ἅ 1F0D	Ἆ 1F0E	Ἇ 1F0F	ἐ 1F10	ἑ 1F11
ἒ 1F12	ἓ 1F13	ἔ 1F14	ἕ 1F15	□ 1F16	□ 1F17	Ἐ 1F18	Ἑ 1F19	Ἒ 1F1A
Ἓ 1F1B	Ἔ 1F1C	Ἕ 1F1D	□ 1F1E	□ 1F1F	ἠ 1F20	ἡ 1F21	ἢ F22	ἣ 1F23
ἤ 1F24	ἥ 1F25	ἦ 1F26	ἧ 1F27	Ἠ 1F28	Ἡ 1F29	Ἢ 1F2A	Ἣ 1F2B	Ἤ 1F2C
Ἥ 1F2D	Ἦ 1F2E	Ἧ 1F2F	ἰ 1F30	ἱ 1F31	ἲ 1F32	ἳ 1F33	ἴ 1F34	ἵ 1F35
ἶ 1F36	ἷ 1F37	Ἰ 1F38	Ἱ 1F39	Ἲ 1F3A	Ἳ 1F3B	Ἴ 1F3C	Ἵ 1F3D	Ἶ 1F3E
Ἷ 1F3F	ὀ 1F40	ὁ 1F41	ὂ 1F42	ὃ 1F43	ὄ 1F44	ὅ 1F45	□ 1F46	□ 1F47
Ὀ 1F48	Ὁ 1F49	Ὂ F4A	Ὃ 1F4B	Ὄ 1F4C	Ὅ 1F4D	□ 1F4E	□ 1F4F	ὐ 1F50
ὑ 1F51	ὒ 1F52	ὓ 1F53	ὔ 1F54	ὕ 1F55	ὖ 1F56	ὗ 1F57	□ 1F58	Ὑ 1F59
□ 1F5A	Ὓ 1F5B	□ 1F5C						

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 06, June-2023

B. Ede language encoding

The Ede alphabet is also classified into the Latin family, with 76 Ede characters including uppercase and lowercase characters as shown in Table II [**Error! Reference source not found.**], [[6]]. Of which 68 characters are the basic components of almost all Unicode fonts, and 8 characters (ĕ, ŏ, ŏ, ŭ, Ĕ
included in the Unicode encoding. [**Error! Reference source not found.**].

TABLE II. EDE ALPHABET

Consonant	***Uppercase***	B	Ɓ	Č	D	Đ	G	H
		J	K	L	M	N	Ñ	P
		R	S	T	W	Y		
	Lowercase	b	ɓ	č	d	đ	g	h
		j	k	l	m	n	ñ	p
		r	s	t	w	y		
Vowel	***Uppercase***	A	Ă	Â	E	Ĕ	Ê	
		I	Ĭ	O	Ŏ	Ô		
				U	Ŭ			
	Lowercase	a	ă	â	e	ĕ	ê	
		i	ĭ	o	ŏ	ô		
		ơ	ŏ	u	ŭ	ư	ŭ	

Where:

Uppercase

Lowercase

Each letter of the Ede language is also mapped into a continuous region in the Unicode encoding. The area selected for mapping has a range from1F00:1F25. The basis for choosing this region is because it is a continuous region containing characters and characters in this region do not appear in Ede documents.

Unlike Vietnamese letters, Ede letters must be converted to a combination code with two characters before being encoded, so that they can be considered as one character when sorted. The rules for converting letters (ĕ, ŏ, ŏ, ŭ, Ĕ, Ŏ, Ŏ, Ŭ) to a character are shown in Table III.

TABLE III. REGULATIONS TO CONVERT THE LETTER EDE IN THE FORM OF A COMBINATION CODE INTO 1 CHARACTER.

Ede letter with 2 characters	**Alternative character**
	!
	@
ŏ	#
ŭ	$

The mapping of Ede letters and corresponding conversion characters to the extended Greek character area is shown in Table IV.

TABLE IV. MAPPING THE LETTER EDE INTO THE EXTENDED GREEK CHARACTER AREA

EDE LANGUAGE CHARACTERS								
a	ă	â	b	ɓ	č	d	đ	e
ĕ	ê	!	g	h	i	ĭ	j	k
l	m	n	ñ	o	ŏ	ô	@	ơ
#	p	r	s	t	u	ŭ	ư	$
w	y							

Extended greek character area								
ἀ 1F00	ἁ 1F01	ἂ 1F02	ἃ 1F03	ἄ 1F04	ἅ 1F05	ἆ 1F06	ἇ 1F07	Ἀ 1F08
Ἁ 1F09	Ἂ 1F0A	Ἃ 1F0B	Ἄ 1F0C	Ἅ 1F0D	Ἆ 1F0E	Ἇ 1F0F	ἐ 1F10	ἑ 1F11
ἒ 1F12	ἓ 1F13	ἔ 1F14	ἕ 1F15	□ 1F16	□ 1F17	Ἐ 1F18	Ἑ 1F19	Ἒ 1F1A
Ἓ 1F1B	Ἔ 1F1C	Ἕ 1F1D	□ 1F1E	□ 1F1F	ἠ 1F20	ἡ 1F21	ἢ F22	ἣ 1F23
ἤ 1F24	ἥ 1F25							

III. EXPERIMENTAL SORTING OF WORD ITEMS IN THE LEXICAL DATABASE

In order to arrange the items in alphabetical order, we experiment with four basic sorting methods: bubble sort; insertion sort; sort select; quick sort [5], to select the sorting method used to sort the items in the lexicon. Based on the execution time after the experiments, we choose the sorting method to include the solution of sorting items in the Vietnamese-Ede vocabulary database. Experimental results on 4 samples, with 10 times per sample, according to 4 sorting methods, are shown in Table V. The details of the experiments are shown in Table VI.

Through the results of the tests on Vietnamese and Ede samples in Table 5, this is the basis for the paper to choose the quick sort method as the sorting method for the array containing the items after being encoded.

TABLE V. TEST RESULTS BY 4 SORTING METHODS

pattern	**Number of experiments**	**Execution time (‰ second)**			
		Bubble sort	**Sort select directly**	**Insert sort**	**Quick sort**
9.297 Ede entries	10	0:0:02.820	0:0:01.479	0:0:00.657	0:0:00.106
17.968 Ede entries	10	0:0:09.477	0:0:04.315	0:0:04.240	0:0:0.188
11.358 Vietnamese entries	10	0:0:02.290	0:0:02.286	0:0:00.268	0:0:00.265
34.375 Vietnamese entries	10	0:1:14.227	0:0:02.286	0:0:13.450	0:0:00.760

TABLE VI. DETAILS OF ATTEMPTS WITH 4 SORTING METHODS

Pattern	**Number of tries**	**Execution time (‰ second)**			
		Bubble sort	**Sort select directly**	**Insert sort**	**Quick sort**
9.297 Ede entries	1	0:0:02.952	0:0:01.492	0:0:00.603	0:0:00.100
	2	0:0:02.961	0:0:01.510	0:0:00.664	0:0:00.099
	3	0:0:02.783	0:0:01.500	0:0:00.595	0:0:00.103
	4	0:0:02.901	0:0:01.479	0:0:00.624	0:0:00.111
	5	0:0:02.696	0:0:01.495	0:0:00.631	0:0:00.110
	6	0:0:02.705	0:0:01.450	0:0:00.587	0:0:00.104
	7	0:0:02.670	0:0:01.540	0:0:00.715	0:0:00.099
	8	0:0:03.008	0:0:01.483	0:0:00.703	0:0:00.111
	9	0:0:02.725	0:0:01.423	0:0:00.723	0:0:00.110

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 06, June-2023

	10	0:0:02.804	0:0:01.414	0:0:00.730	0:0:00.117
	Average	**0:0:02.820**	**0:0:01.479**	**0:0:00.657**	**0:0:00.106**

17.968 Ede entries	1	0:0:09.925	0:0:04.484	0:0:04.829	0:0:0.162
	2	0:0:08.757	0:0:04.420	0:0:04.807	0:0:0.163
	3	0:0:08.539	0:0:04.699	0:0:03.490	0:0:0.207
	4	0:0:09.811	0:0:05.045	0:0:03.802	0:0:0.196
	5	0:0:09.371	0:0:03.874	0:0:03.725	0:0:0.165
	6	0:0:10.452	0:0:03.900	0:0:04.463	0:0:0.162
	7	0:0:09.145	0:0:04.124	0:0:04.845	0:0:0.199
	8	0:0:09.067	0:0:03.889	0:0:04.876	0:0:0.197
	9	0:0:10.217	0:0:04.405	0:0:03.741	0:0:0.230
	10	0:0:09.487	0:0:04.318	0:0:03.829	0:0:0.205
Average		***0:0:02.290***	***0:0:02.286***	***0:0:00.268***	***0:0:00.265***
11.358 Vietna_ mese entries	1	0:0:02.046	0:0:01.920	0:0:00.239	0:0:0.340
	2	0:0:02.028	0:0:02.091	0:0:00.247	0:0:0.250
	3	0:0:02.511	0:0:02.300	0:0:00.309	0:0:0.225
	4	0:0:02.542	0:0:02.165	0:0:00.330	0:0:0.240
	5	0:0:01.918	0:0:01.991	0:0:00.235	0:0:0.234
	6	0:0:02.090	0:0:02.741	0:0:00.257	0:0:0.286
	7	0:0:02.418	0:0:02.314	0:0:00.235	0:0:0.220
	8	0:0:02.433	0:0:02.223	0:0:00.343	0:0:0.350
	9	0:0:02.345	0:0:02.870	0:0:00.252	0:0:0.241
	10	0:0:02.576	0:0:02.240	0:0:00.232	0:0:0.267
34.375 Vietna_ mese entries	1	0:1:14.166	0:0:18.910	0:0:12.168	0:0:0.795
	2	0:1:13.985	0:0:19.425	0:0:13.462	0:0:0.686
	3	0:1:14.374	0:0:17.841	0:0:14.679	0:0:0.826
	4	0:1:13.956	0:0:19.410	0:0:14.835	0:0:0.748
	5	0:1:14.126	0:0:21.091	0:0:12.963	0:0:0.795
	6	0:1:14.212	0:0:17.862	0:0:12.651	0:0:0.875
	7	0:1:14.028	0:0:20.420	0:0:13.806	0:0:0.842
	8	0:1:13.825	0:0:18.798	0:0:13.868	0:0:0.592
	9	0:1:15.006	0:0:18.688	0:0:12.731	0:0:0.717
	10	0:1:14.589	0:0:19.983	0:0:13.338	0:0:0.733
Average		***0:1:14.227***	***0:0:02.286***	***0:0:13.450***	***0:0:00.760***

IV. EXPERIMENTAL RESULTS

Currently, if with the Order by clause of the query statement, the results are sorted in alphabetical order of Vietnamese string attributes. With the accented alphabetic the letter đ, there is no result of alphabetical arrangement in Vietnamese. The result when using the Order by clause in the SQL statement is shown in Figures 1 and 2. Figure 1 shows the results when executing the query Select Viet From VIET Order by Viet. Figure 2 shows the results when executing the query Select Viet From VIET Order by CS_SX. The CS_SX attribute is an added attribute according to the above solution for sorting items in the Viet-Ede datastore.

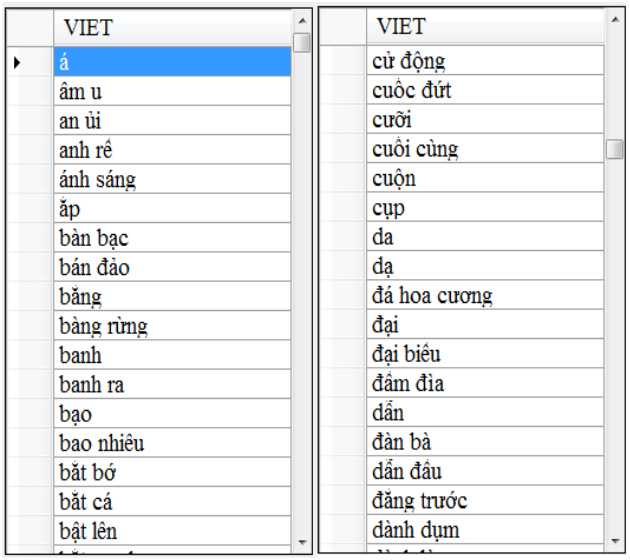


Fig 1. Result of executing sort command with Vietnamese entries

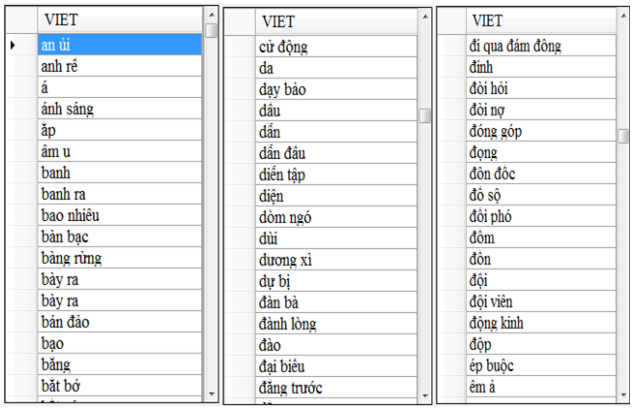


Fig 2. Result *of executing the command to sort Vietnamese entries with the sort index when coded*

With Ede letters, the same situation is encountered in Vietnamese. In addition, Ede language also has a case of handling letters that are combined in the form of a combination code. The results when using the Order by clause in the SQL statement are shown in Figures 3 and 4. Figure 3 shows the result when executing the *query Select Ede From EDE Order by Ede*. Figure 4 shows the result when executing the query *Select Ede From EDE Order by CS_SX*. The CS_SX attribute is an added attribute according to the above solution for sorting entries in the Viet-Ede datastore.

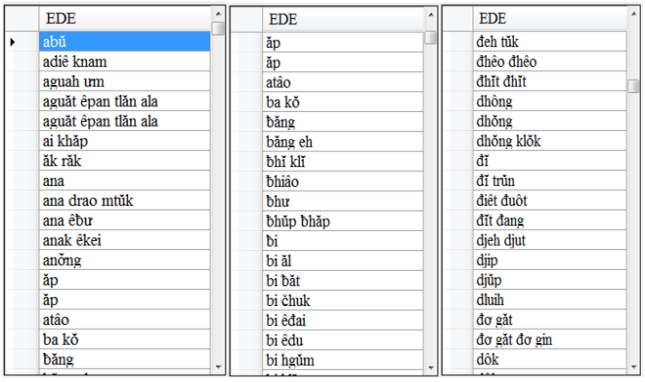
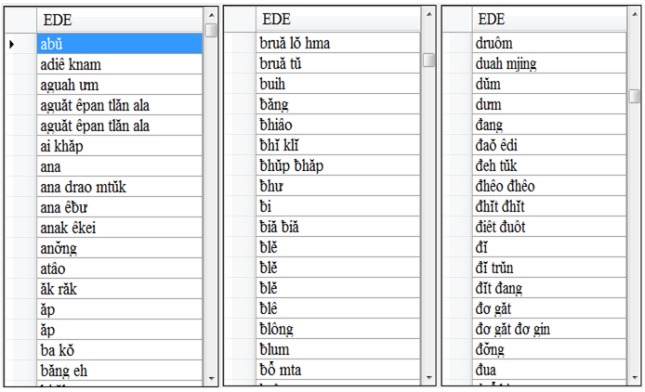


Fig 3. The result of executing the sort command with Ede entries

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 12 Issue 06, June-2023



EDE
abŭ
adiê knam
aguah ưm
aguăt êpan tlăn ala
aguăt êpan tlăn ala
ai khăp
ana
ana drao mtŭk
ana êbư
anak êkei
anŏng
atâo
ăk răk
ăp
ăp
ba kŏ
băng eh

EDE
bruă lŏ hma
bruă tŭ
buih
băng
bhiâo
bhĭ klĭ
bhŭp bhăp
bhư
bi
bia bia
blĕ
blĕ
blĕ
blĕ
blông
blum
bŏ mta

EDE
druôm
duah mjing
dŭm
dưm
đang
đaŏ êdi
đeh tŭk
đhêo đhêo
đhĭt đhĭt
điêt đuôt
đĭ
đĭ trŭn
đĭt đang
đơ găt
đơ găt đơ gin
đŏng
đua

Fig 4. The result of executing the command to sort items from the Ede entries with the sorting index when encrypted

V. CONCLUSION

The solution to sort items in the Vietnamese-Ede bilingual vocabulary database has been sorted on the attribute containing Vietnamese entries and Ede entries. The results are sorted according to the Vietnamese and Ede alphabetical order when using the Order by clause in the SQL query statement in the Viet-Ede datastore.

This solution contributes to solving the problem of arranging Vietnamese entries and Ede entries in the Vietnamese-Ede bilingual vocabulary database in alphabetical order in the data query statement with data sorting.

In the next orientation, the paper will apply this solution to integrate into applica-tions that edit tables such as Winword, Excel to arrange columns or rows in Ede data tables.

REFERENCES

[1] Doan Van Phuc: *Ede phonetics*, Social science, Hà Nội, 1996.

[2] Hoang Thi My Le, Vilavong Souksan, Phan Huy Khanh: *Using Unicode in Encoding the Vietnamese Ethnic Minority Languages, Applying for the EDe Language*, Proceeding of the International Conference on Knowledge and System Engineering, KSE 2013, HaNoi, pp. 137-148, 2013.

[3] Hoang Thi My Le Phan Huy Khanh: *The solution to build a Vietnamese-Ede bilingual vocabulary database based on the Vietnamese-Ede interaction model*, No 5 (2), pp. 36–40, 2017.

[4] Le Hoang Thi My, Khanh Phan Huy: *Deploying environment for processing Ede ethnic minority language in Vietnam*, IEEE International Conference on System Science and Engineering (ICSSE), 2017..

[5] Robert Sedgewich: *Algorithm*, NXBKH & KT, 2003, https://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-840.pdf

[6] Department of Education and Training DakLak: *Ede Grammar*, Education publisher, 2011.