World of Computer Science and Information Technology Journal (WCSIT)
ISSN: 2221-0741
Vol. 6, No. 1, 1-7, 2016

New Automatic Search and Update Algorithms of Vietnamese Abbreviations

Nguyen Nho Tuy
Vietnam Posts and Telecommunications Group
(VNPT) Danang Branch
Danang City, Vietnam

Phan Huy Khanh
University of Science and Technology
- The University of Danang
Danang City, Vietnam

Abstract- Abbreviations in documents are widely used in various fields and in many languages including Vietnamese. In fact, currently, abbreviations are regularly repeated and unclearly used, demand for abbreviation use is increasing, which requests a plentiful source of abbreviations which is conveniently saved and used, easily updated and consistently exploited. In this article, we propose some abbreviation search algorithms on the Internet in order to automatically update into database of Vietnamese abbreviations for many purposes during language processing and database exploitation.

Keywords- abbreviation; acronym; database; abbreviation searching programs; automatic search Vietnamese abbreviations.

I. INTRODUTION

Abbreviations are familiar in daily life and have been widely used in almost the written language system in the world so far, including Vietnamese. In newspapers, magazines, we often see common abbreviations such as TƯ (Trung ương), UBND (Uỷ ban nhân dân), and also English abbreviations such as WTO (World Trade Organization, etc. Thanks to abbreviations, all texts are shorter and simpler while express more capacity of information. The fact that abbreviations are often used makes the abbreviation system increasingly diversified and abundant. On the one hand, users (NSD) have many abbreviations to choose and use, on the other hand, the users also run into a lot of difficulty in finding, searching its meanings and proper using of such abbreviations.

With regard to abbreviations, there are some dictionaries today such as Dictionary of telecommunication, Dictionary of abbreviations in telecommunication [8]; websites of Abbreviations, but mainly in foreign languages. The need of abbreviations is higher and higher, wider and wider and indispensable, especially brands, trademarks, etc.

Contents of this article include: Firstly, we present information about abbreviations, history of abbreviation development, principles to create abbreviations, classification of abbreviations and influential factors in abbreviation generation. Next, we present database of abbreviations, algorithms and programs of automatic abbreviation collection on the Internet, statistically assess results and give solution of abbreviations. The final part is conclusions.

II. INFORMATION ABOUT ABBREVIATIONS

2.1. Definition and terms

The term "“chữ viết tắt” (In English: abbreviation) has not appeared in Common Vietnamese dictionary in current market1 including in “Từ điển Bách khoa Việt Nam” Vol. 1 (Letters A-Đ2), however, it is very familiar in daily life.

We often see abbreviations or acronyms. They are used for generating abbreviations that are different from common written languages; abbreviations are used when we have to repeatedly write a word, phrase, sentence or paragraph for convenience [8]. For a long time ago, people used abbreviations to inscribe on stone, wood, etc in order to save time, force and material. According to Manuel Zahariev[14], abbreviations are originated in Ancient Greek, *acronym* includes *akron* (the last or first one) and *onoma* (name or voice). According to some English dictionaries, abbreviations are the way to form new shorter words by using initial letters, or last letters or any letters of a word. For example, UNESCO stands for “United Nations Educational, Scientific and Cultural Organization”, etc.

We also see abbreviations in short form, it means that a

[1] Vietnamese-English dictionary, Bui Phung, published by global publishing house in 1998.

[2] Vietnam encyclopedia compilation steering council compiled. Vietnam encyclopedia compilation center published in 1995.

phrase or a paragraph have some characters abridged or have one part extracted, chosen or replaced to form a set of new characters, in order that writing and saying are more convenient. For example, Vietnamese abbreviations are used for geographical areas, for example, Thanh Land, Nghe Land, Quang Land, etc.

In the progress of Internet explosion, generally, written languages have been developed towards a new direction thanks to the use of various abbreviations and conventional signs. For example, in English, email, messages, IMHO stands for "in my humble opinion", comic signs ☺, ☹, U (you), etc. The use of abbreviations in fields of information technology and communication today on one hand makes users beneficial, on the other hand, such diversity or abuse of abbreviations also troubles the users.

2.2. History of abbreviations

Abbreviations have been widely used for a long time ago in foreign countries. For example, SPQR stands for "Senatus Populusque Romæ" and has appeared for nearly 2000 years [14], QED stands for "Quod Erat Demonstrandum" (proved) in "Ethica More Geometrico Demonstrata" of a philosopher, Benedictus de Spinoza (1632-1677).

In Vietnam, today, there are some researches into Vietnamese abbreviations [4][12], however, such researches are not complete and systematic, although Vietnamese abbreviations have been early formed. The formation of "chữ Nôm" (an ancient ideographic vernacular script of the Vietnamese language) since the 18th century has been other way to write "chữ Hán" (Chinese writing), replace "chữ Hán" after nearly one thousand years of occupation and colonization by the Han [2][3]. In the "chữ Nôm", each "chữ Nôm" is square, is formed by putting "chữ Hán" together in the form of onomatopoeia, pictographic or reducing characters, abbreviation. For example, the Chinese writing 共 (total) is reduced its characters into "chữ Nôm" 𠔩 (khạng), "chữ Hán" 爲 (vi) is reduced its characters into "chữ Nôm" 爫 (làm).

When Vietnamese national language (Current Vietnamese language) had been widely used, abbreviations have been used. The pen name C.D. standing for Chương Dân is official name of Phan Khôi in "Đông Pháp Thời Báo" in 1928. Today, Vietnamese abbreviations are being used increasingly widely in many fields.

Many authors think that Vietnamese abbreviations refer to a grammar [1][9][10]. According to Prof. Nguyen Tai Can, we "*use abbreviation in form of one syllable rather than in form of initial letters. The acronyms such as DT (danh từ), VN (Việt Nam), HTX (hợp tác xã), etc only are used in writing documents".* Although there are many views of the use of abbreviations, abbreviations are existing as an internal part of Vietnamese language, and there are many abbreviation applications in communication, text processing, data exploitation [5], etc.

2.3. Principles of abbreviation generation

Based on the results of analysis, the demand and current status of the abbreviation use in daily life, we proposed 07 Principles of abbreviation generation as detailed[4], and now, we supplement 2 new Principles of abbreviation generation (Principles 8 and 9).

1. 7 Principles of abbreviation generation that have been developed: Principle of abbreviation; principle of word connection; principle of short connection by meaningful words; principle of sub-letters; principle of connection of foreign languages; principle of borrowing of abbreviations in foreign languages; principle of random abbreviation.

2. 2 new principles include:

   1) Principle of encrypted abbreviations:

In many fields and sections, reminiscent abbreviations are used in conformity with a predefined rule to encrypt the phrase. All encrypted abbreviations often must satisfy:

i. Encrypted abbreviations are often issued by an organization with scope of use and application.

ii. Encrypted abbreviations are unique and unduplicated to avoid ambiguity.

iii. Encrypted abbreviations have often new characters used according to a predefined rule.

For example, lists and tables in database, list of national codes, regional codes, sectional codes, and codes of telecom fiber optic cables, etc.

   2) Principle of abbreviations in database:

According to studying in theories of searching problems, relevant practical results and the efficient use of abbreviations, we propose some principles of applying index abbreviations in order to search data in large database:

i. Abbreviations only used English letters (not Vietnamese words) and digits 0…9

ii. Don't use special characters: punctuation marks, space (SP)

iii. Abbreviations are reminiscent, short, not unduplicated, and not unclear: Users immediately image abbreviations after determining request for information searching.

iv. Implement index of database on the established fields of abbreviations.

2.4. Influential factors in the new abbreviation generation

According to the field survey, we propose 4 influential factors in the generation of new abbreviations, particularly:

Number of characters: Abbreviations shall not be too long. In general, common number of characters of an abbreviation should not be more than 18 characters.

Marks in Vietnamese language: Avoid vowel with mark such as *â, ă, ơ, ê, etc*; don't use grave, acute, question mark and dot below in abbreviations in order to avoid misunderstanding, difficulties in speaking.

Spiritual factors for East Asians: Select number of characters of an abbreviation. Avoid number 2, number 4 or avoid number of characters of an abbreviation according to the conception of "*birth, old age, illness, death"*. In order to generate the word "*birth*", the number of characters of an abbreviation shall be 5, 9, 13, etc, and in order to generate the word "*old age*", the number of characters of an abbreviation shall be 2, 6, 14, etc.

Syllable: Select abbreviations so that when being read, such abbreviations form opening and deep hollow notes. People often choose *a*, *ô*, *i*, or *ex*, *ec*, rather than *ê*, *ơ*.

Two last factors often are specially considered when finding abbreviated name of enterprises, companies, brands, trademarks, organizations, projects, etc.

2.5. The use of abbreviations

Generally, users shall define or explain all abbreviations in documents. There are two cases as below:

Using available abbreviations: Abbreviations are defined and explained previously, or commonly used, not unclear.

Using new abbreviations: Defining and using abbreviations right after initial appearance in documents in the form of:

<Complete phrase > (<Abbreviation>)

The above principles of abbreviation generation allow us to refer 05 signs of abbreviations in a Vietnamese document, particularly:

1) Abbreviations are placed in brackets (..), or placed after the phrases: "viết tắt là", "viết tắt", "gọi tắt là"…(hereinafter referred to as…, hereinafter called, etc.) when the abbreviations are defined initially.

2) Abbreviations are capital letters (lowercase in normal letters)

3) Abbreviations include special letters or marks: *and (&), cross mark (/), dash (-), dot (.), space,* and letters and digits, etc.

4) Abbreviations are words whose number of characters may be 18.

5) Vietnamese abbreviations shall not include vowels *â, ă, ơ, ê, ô…* don't use marks such as *grave, acute, question mark and dot below.*

2.6. Ambiguity of abbreviations

Ambiguity of abbreviations is not rare, the ambiguity is formed by natures: difficulty in understanding abbreviations, arbitrary abbreviations, not complying with rules, difficulty in defining the meaning of abbreviations:

For example: VH: Văn hóa, Văn học; Abbreviations are local, uncommon: Cao Xà Lá: Cao su, Xà phòng, Thuốc lá; Phối kết hợp: Phối hợp, kết hợp; not complying with rules: SKZ: **s**úng **k**hông giật/**z** ...

The principles of abbreviation generation 1 – 8 often cause Ambiguity. The principles 8, 9 do not cause Ambiguity within scope and application of abbreviations. However, Vietnamese abbreviations in general have the following characteristics:

III. Difficulty in defining the meaning of abbreviations due to the way of writing

IV. Abbreviations are often formed to be easy to speak, to remember and convenient, thus abbreviations are often concise and polysemous.

V. Abbreviations continuously change; the formation of language @ and use of foreign language abbreviations make abbreviations increasingly plentiful and diversified;

2.7. Update of abbreviations

In the world, there are many researches into abbreviations and issues of database establishment, abbreviation update by manual, online and automatic methods. Manuel Zahariev [14] studied the process of automatic formation and generation of English abbreviations. Abbreviation dictionary website [20] saves more than 5 million abbreviations in multiple languages, mostly updated by manual method; however, some solutions of online abbreviation update are given to the users in available form, abbreviations formation is advised and duplication is warned; and then edited and put into the database. There are further in-depth studies in the update, search and extension of Chinese abbreviations in software maintenance [15]. The researches by authors David Sánchez and David Isern launched a method of automatic update and search without supervision for English abbreviation generation, extraction of definitions with abbreviations from the Website, a new approach to establish the abbreviation archive rather than manual method by AcronymFinder [15] [16].

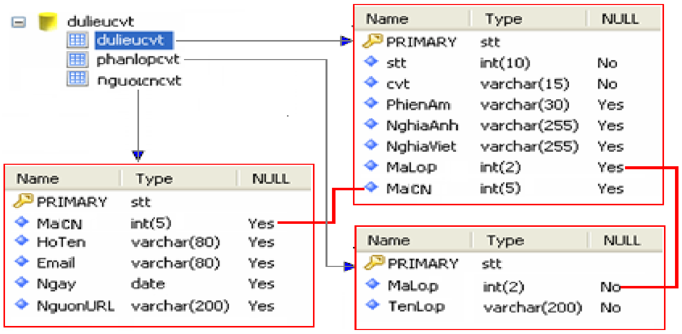
In Vietnam, there is almost no in-depth study in Vietnamese abbreviations; no adequate attention to abbreviation update and saving. However, the initial studies were implemented in incoherent, separate manner [5] [12]; but automatic update on the Internet are less mentioned and has no remarkable results.

VI. DEVELOPING ALGORITHMS AND PROGRAMS OF AUTOMATIC UPDATE OF VIETNAMESE ABBREVIATIONS

3.1. Developing Model of database

Classification of abbreviations: There are many methods in classification of abbreviations, basing on field of use, site, etc. In article published in 2006[4], we recognized 9 fields; and by now, with classifications of abbreviations up on field of use, we recognized the 12 main fields (table 1).

We develop database (database) for abbreviations, including 3 tables of DULIEUCVT (data of abbreviation), PHANLOPCVT (classification of abbreviation) and NGUOICNCVT (editor of abbreviation) with relations as figure below.



Name	Type	NULL
PRIMARY	stt	
stt	int(10)	No
cvt	varchar(15)	No
PhienAm	varchar(30)	Yes
NghiaAnh	varchar(255)	Yes
NghiaViet	varchar(255)	Yes
MaLop	int(2)	Yes
MaCN	int(5)	Yes

Name	Type	NULL
PRIMARY	stt	
MaCN	int(5)	Yes
HoTen	varchar(80)	Yes
Email	varchar(80)	Yes
Ngay	date	Yes
NguonURL	varchar(200)	Yes

Name	Type	NULL
PRIMARY	stt	
MaLop	int(2)	No
TenLop	varchar(200)	No

Figure 1. Relations of database of abbreviations.

Table DULIEUCVT contains abbreviation information including: order of abbreviations, field of abbreviations, phonetic field to easily read the fields, field of meanings (explanations) in English and fields in Vietnamese, fields of layer codes and fields of updated codes which are outer locks connecting to two databases accordingly. Table DULIEUCVT contains all possible abbreviations for exploitation and continuous update. Table PHANLOPCVT enlists layers of abbreviations including code and name of layer.

3.2. Proposals of abbreviation auto-update algorithms from the Internet

We use different sources of abbreviations to update into the database. The update process is conducted manually, directly in Winword documents from many different sources such as books, newspapers, magazines, legal documents, scientific reports or the life, etc. However, the source of abbreviations on the Internet is very abundant. We develop abbreviation auto-update algorithms based on the Internet as follow.

Introduction on theoretical model of Search Engine

Search Engine - abbreviated as SE is a tool established on the web base, allowing users to search for information. Basic components of a search engine, Figure 2.

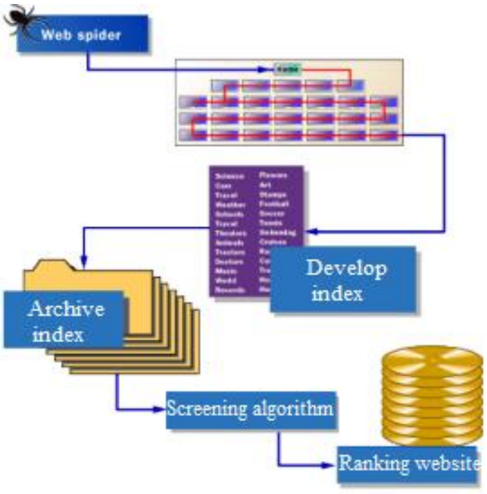


Figure 2. Basic components of a search engine (Image on the internet)

That is called the Spider program that explores the website, performs the scanning function and sets up website index, checks links in the website. All things searched by Spider will be saved in a huge library and divided into the index. And then, "library" will be screened and ranked through hundreds of algorithms. The mechanism of operation and internal algorithms of search engine are mostly located within the security.

Base on the ideal of search engine, we develop the abbreviation search engine whose operating principles were introduced in [13]. The algorithm describes operation of the engine [4] in the Internet environment as follows:

```
 Algorithm: Vietnamese abbreviation auto-
search on the Internet
 Input : Address URL
 Output: Data of abbreviations in table
TUDONGCVT
 Open intermediary database
 Define operative URL
 Save URL in intermediary database
 Activate abbreviation counter
 Repeat
     Open a file HTML
     Read content respectively HTML
     Dissect data (remove space and tags HTML)
     Find abbreviations basing on aware signals
     If found abbreviations Then
         Check whether abbreviations exist or
       not?
         If abbreviations exist Then
               Increase abbreviation counter
         Else
               Save abbreviations and assign the
             corresponding value by 1
               Extract a sentence containing
             abbreviations
         End If
     End If
 Until no more HTML
```

3.3. Setting up the program

The program is set up based on PHP order codes, HTMP cards on the Web: thuthapv5.php according to the detailed source codes [15]. In this article, we only present administrative procedures and use legend by the mark // before or after each sequence of instructions.

```
<html>
// HTML cards
<body>
<?PHP
//-- Function of testing to reject strings //which
are not abbreviations:
function testdauhieucvt($string)
{
  // Test the $string return logic values
   return $dauhieu;
}
//-- Function of extract meanings of abbreviations
in a sentence containing abbreviations
function nghiacau($cauxet,$cvt)
{
return $nghia;
}
//======= Main program:=============
// Connect Database. If successful connecting:
   mysql_select_db("dulieucvt");
// Define URL to be processed:
// --Searching all links on the URL --
// Take links, select links on the HTMP pages such
as the links in the form of.htm|.html|.php|.aspx,
assign to the list $dslienket
// ---- Read each connection in $dslienket and
search abbreviations, save in database
// Loop reading each connection:
for($ii = 0; $ii < count($dslienket); $ii++)
{
 // Open connection
 $fd = fopen($url_ii,"r");
 while( $doan = fgets($fd,1024) )  // Loop reading
each paragraph
 {
   //Only consider the paragraph containing mark (...) and
else blank
   $doan = trim($doan);
//Delete blank in the beginning and ending of strings
   // Remove HTML tags:
   $doan = strip_tags($doan);
   // Extract a sentence
   // Loop processing each sentence
   for( $i = 0; $motcau[$i]; $i++ )
      {
      // Only consider the paragraph containing mark
(...) and else blank
      $xetcau = $motcau[$i][$j];
         $btcqdaungoac="/[^\(]+[\)$]/";// RE select
phrases ()
      // $tuduocchon save phrases (...)which may be
abbreviations
        for( $k = 0; $tuduocchon[$k]; $k++ )// Only
analyze the paragraph containing mark (...) and else
blank each phrase
       {
       if testdauhieucvt($tudangxet)))  // Satisfy
identification signs of abbreviations
           {
     // Call function extracting meanings of abbreviations
         $nghiacvt=nghiacau($xetcau,$tudangxet);
         // Check whether abbreviations exist or not?
             // Save $tudangxet,"$nghiacvt, $xetcau,
$doan
          // Open database ‘tttdviet’
           } //if
          } //for k
    } // for i
 } // While
fclose($fd);
} // For ii
?>
</body>
</html>
```

The above program uses regular expressions and functions on PHP in order to process strings with regular expressions.

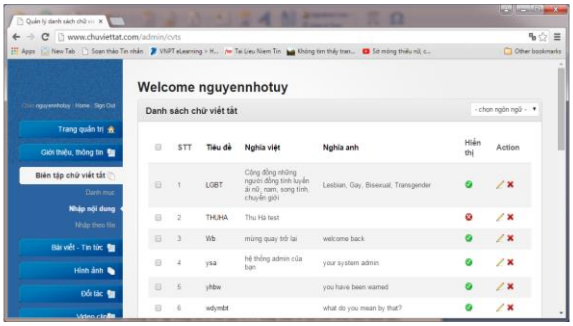
For example, `$btcqdaungoac="/[^\(]+[\)$]/"`

is an official expression selecting the string on the quotation marks;

The function: `preg_match_all($btcqdaungoac,$xetcau,$Upwords)` extracts string on the quotation marks from the current sentence which is being considered to save in the two-way variable `$Upwords`.

After searching, it is necessary for the participation of experts in editing and correcting data. The updating process will include test and warnings for repeat of abbreviations or repeat in meanings.

Interface of Admin website for update and edit of abbreviations will be developed as the figure 3.



Quản lý danh sách chữ vi...

www.chuviettat.com/admin/cvts

Welcome nguyennhotuy

Chào nguyennhotuy | Home | Sign Out

Danh sách chữ viết tắt

- chọn ngôn ngữ -

	STT	Tiêu đề	Nghĩa việt	Nghĩa anh	Hiển thị	Action
□	1	LGBT	Cộng đồng những người đồng tính luyến ái nữ, nam, song tính, chuyển giới	Lesbian, Gay, Bisexual, Transgender	✓	✎ ✗
□	2	THUHA	Thu Hà test		✗	✎ ✗
□	3	Wb	mừng quay trở lại	welcome back	✓	✎ ✗
□	4	ysa	hệ thống admin của bạn	your system admin	✓	✎ ✗
□	5	yhbw		you have been warned	✓	✎ ✗
□	6	wdymbt		what do you mean by that?	✓	✎ ✗

Figure 2. Interface of Admin website for edit and update of database of abbreviations.

3.4. *Statistics of the results and applications*

Thanks to the auto-update of abbreviations and careful editing contents, by now, we have enlisted the number of existing English and Vietnamese abbreviations in database as follows:

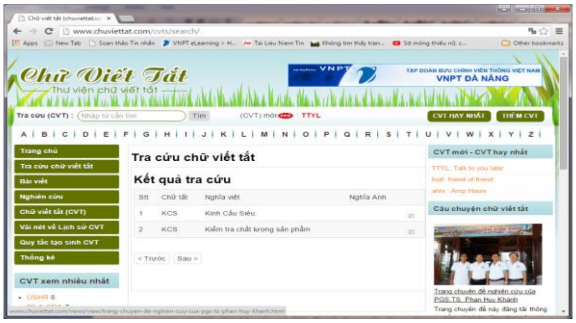
TABLE 1. STATISTICS OF UPDATE OF ABBREVIATION DATABASE

Cate-gory	Fields of abbreviations	Manual update	Auto-matic update	Total	% of auto-update
1	Information technology and communication	754	350	1104	32%
2	Government, political and social organizations	301	120	421	29%
3	Science, technology and engineering	273	253	526	48%
4	Military	202	120	322	37%
5	Medicine	253	255	508	50%
6	Education	301	2378	2679	89%
7	Finance, trade	403	140	543	26%
8	Environmental resources	163	130	293	44%
9	Community communication	121	125	246	51%
10	Religion	0	150	150	100%
11	Proper name	0	75	75	100%
12	Other	0	120	120	100%
	Total	**2771**	**4216**	**6987**	**60%**

According to the statistics result, auto-update achieved 60% although much data of abbreviations is barely updated; abbreviations continuously change. Particularly, education field owns lots of abbreviations, mainly relating to code of colleges, professionals and specialties.

Some applications of exploiting database of abbreviations

We establish a website www.chuviettat.com (fig. 3) containing database of abbreviations and managing online search of abbreviations in Vietnamese and English to serve users intensively.



Chữ viết tắt (chuviettat.co...

www.chuviettat.com/cvts/search/

Chữ Viết Tắt

Thư viện chữ viết tắt

TẬP ĐOÀN BƯU CHÍNH VIỄN THÔNG VIỆT NAM
VNPT ĐÀ NẴNG

Tra cứu (CVT) : Nhập từ cần tìm [Tìm] (CVT) mới : TTYL [CVT HAY NHẤT] [THÊM CVT]

A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z |

Tra cứu chữ viết tắt

Kết quả tra cứu

Stt	Chữ tắt	Nghĩa việt	Nghĩa Anh
1	KCS	Kinh Cầu Siêu	
2	KCS	Kiểm tra chất lượng sản phẩm	

< Trước Sau >

CVT mới - CVT hay nhất

TTYL: Talk to you later
fof: friend of friend
ahrs : Amp Hours

Câu chuyện chữ viết tắt

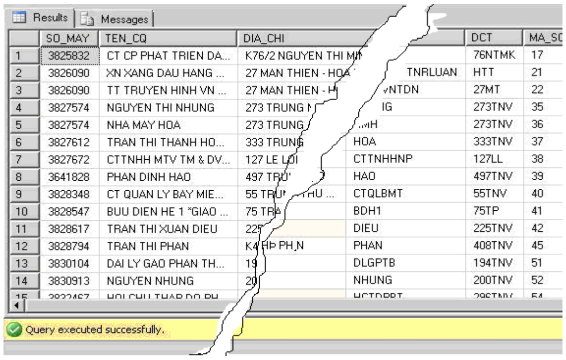
Trang chuyên đề nghiên cứu của PGS.TS. Phan Huy Khánh

Trang chuyên đề này đăng tải thông

Figure 3. Interface of website for abbreviation exploitation.

Abbreviation application in database exploitation: Capacity of the information search depends on not only the resource capacity of the system or searching algorithm but also operative and processing time on users' computer (users).

From the access, study, update and formation of abbreviation database, we use abbreviations in practical works. We announced a solution by developing a generation function for abbreviations (abbreviations) to apply into the re-establishment of database (database) upon the customer information at Switchboard 108 VNPT Da Nang. We also apply the practicality of the solution like index abbreviations, and the short insert of the abbreviation at abbreviation search does bring practical benefits for Switchboard 108 VNPT in information search among customers [5].



Results | Messages

	SO_MAY	TEN_CQ	DIA_CHI		DCT	MA_SC
1	3825832	CT CP PHAT TRIEN DA...	K76/2 NGUYEN THI MI		76NTMK	17
2	3826090	XN XANG DAU HANG ...	27 MAN THIEN - HOA	TNRLUAN	HTT	21
3	3826090	TT TRUYEN HINH VN ...	27 MAN THIEN - H	VNTDN	27MT	22
4	3827574	NGUYEN THI NHUNG	273 TRUNG N	IG	273TNV	35
5	3827574	NHA MAY HOA	273 TRUNG	MH	273TNV	36
6	3827612	TRAN THI THANH HO...	333 TRUNG	HOA	333TNV	37
7	3827672	CTTNHH MTV TM & DV...	127 LE LOI	CTTNHHNP	127LL	38
8	3641828	PHAN DINH HAO	497 TRU	HAO	497TNV	39
9	3828348	CT QUAN LY BAY MIE...	55 TRU THU ...	CTQLBMT	55TNV	40
10	3828547	BUU DIEN HE 1 "GIAO ...	75 TRA	BDH1	75TP	41
11	3828617	TRAN THI XUAN DIEU	225	DIEU	225TNV	42
12	3828794	TRAN THI PHAN	K4 HP PH,N	PHAN	408TNV	45
13	3830104	DAI LY GAO PHAN TH...	19	DLGPTB	194TNV	51
14	3830913	NGUYEN NHUNG	20	NHUNG	200TNV	52
15	3832467	HOI CHU THAP DO PH		HCTDPBT	296TNV	54

Query executed successfully.

Figure 4. Result of database establishment for phone subscriber lookup through Switchboard 108 VNPT-Da Nang

VII. CONCLUSIONS

Approach and study abbreviations, aggregate the principles of abbreviation generation, build database of abbreviations to serve users in exploitation, storage, statistics and use; especially propose some abbreviation search algorithms on the Internet in order to search new abbreviations, auto-update database of Vietnamese abbreviations for many purposes during processing languages and exploiting database.

The use of abbreviations in establishing indexes of database for better exploitation of special database for searching is meaningful, helps to improve capacity and performance of data exploitation in reality.

In addition, the coherent and universal use of abbreviations is to standardize the system of abbreviations for users, gradually enrich the system of vocabulary and contribute to the development of language. The proposal of rules, methods in management, establishment of an abundant storage, convenient exploitation and use, easy update, formation of forum, new addition of abbreviations, etc. are necessary and beneficial.

We continue to expand the storage of abbreviations in many fields, increase the number of auto-updated abbreviations, evaluate the frequency, frequency of appearance and utilize abbreviations; enhance the transfer into many different languages; and expand the searching capacity in multi-languages like Vietnamese-Kinh, language of ethnic minorities (Cham, Ede, Thai, Kh'me, etc.), English, French, Chinese, etc. This is a righteously oriented pathway to satisfy a common interest.

ACKNOWLEDGMENT

Author group sincerely send our gratefulness to staffs of Switchboard 1080 VNPT Da Nang to create favorable conditions during the process of approach, establishment of database of abbreviations and abbreviation use, also to exploit and contribute actively for the completion of database.

REFERENCES

[1] Nguyen Tai Can. Vietnamese grammar. Publishing house of university and professional secondary school, Hanoi 1981.

[2] La Minh Hang. Nom in context of regional culture. International conference about Nom, between 12-13/11/2004, national library of Vietnam.

[3] Ngo Thanh Nhan, Ngo Trung Viet and Nom Na group. Nom Na process. Summer conference 2002 at Maine University.

[4] Phan Huy Khanh, Nguyen Nho Tuy. Study to build up database of abbreviations in service 1080 of Da Nang Post office. Summary record of national scientific conference "Some selected issues of information technology and media", 2006.

[5] Phan Huy Khanh, Nguyen Nho Tuy. Abbreviation use in service exploitation of Switchboard 108 VNPT Da Nang City. IJISET (International Journal of Innovative Science, Engineering & Technology), Vol. 3 Issue 1, January 2016, p.222-227.

[6] Phan Huy Khanh. Build database of multi-language vocabulary in form of document RTF Winword. Summary record of national scientific conference ICT. rda2003, page 103-110.

[7] Phan Huy Khanh, Use programming tools macro VBA, build up text processing facilities. Summary record of the third scientific conference, Da Nang University 11/2004, page 255-261.

[8] Nguyen Thanh Viet, Do Kim Bang. Terms in telecommunication abbreviations. Publisher of post office, 1999.

[9] Nguyen Thi Thu Thuy. Vietnamese vocabulary. Remote training curriculum of Can Tho University.

[10] Chim Van Be. Vietnamese grammar. Remote training curriculum of Can Tho University.

[11] Nguyen Thi Thu Thuy, Nguyen Huu Chinh Overview of language and linguistics. Remote training curriculum of Can Tho University.

[12] Huynh Cong Phap, Nguyen Van Hue (2014). Study, collect and build up database of abbreviations in Vietnamese. Journal of Science and Technology of Da Nang University. No 7(80).

[13] Hoang Hiep, Build up searching tools by PHP and MySQL. Journal of Posts and Telecommunications and Information Technology, series 2, 9/2004.

[14] Doctor Manuel Zahariev. Acronyms. Simon Fraser University, Jun 2004.

[15] David Sánchez, David Isern. Seeking Acronym Definitions:a Web-based Approach. Universitat Rovira i Virgili Departament d'Enginyeria Informàtica i Matemàtiques ITAKA Research Group. January 2009

[16] David Sánchez, David Isern. Automatic extraction of acronym definitions from theWeb. Appl Intell (2011) 34. September 2009

[17] Zachary P. Fry. IMPROVING AUTOMATIC ABBREVIATION EXPANSION WITHIN SOURCE CODE TO AID IN PROGRAM SEARCH TOOLS. University of Delaware, 2008

[18] http://www.chuviettat.com/news/view/ma-nguon-chuong-trinh-may-tim-kiem-chu-viet-tat-tren-internet.html

[19] http://chuvietnhanh.sourceforge.net/

[20] http://www.acronymfinder.com

AUTHORS PROFILE

1. Full name: Nguyen Nho Tuy

Qualifications: Master of Computer Science; now: doctoral student research

Unit: Vietnam Posts and Telecommunications Group (VNPT) Danang Branch

Danang City, Vietnam.

2. Full name: Phan Huy Khanh

Qualifications: Associate Professor Ph.D; Natural language processing, artificial intelligence, information systems

Unit: University of Science and Technology - The University of Danang

Danang City, Vietnam.