

A Study on the Audio Comparison Expletive Filtering System using TTS

Won-Cheol Park¹, Koo-Rack Park^{1*}, Jae-Woong Kim¹, Young-Suk Chung¹ and Dong-Hyun Kim²

¹Department of Computer Engineering, Kongju National University, Republic of Korea; pwcfrog@nate.com, ecgrpark@kongju.ac.kr, jykim@kongju.ac.kr, merope@kongju.ac.kr,

²Department of IT Convergence, Woosong University, Republic of Korea; dhkim@wsu.ac.kr

Abstract

Objectives: The use of expletives on online is increasing more and more. As online is used by various age groups, preventing the use of expletives is required. The purpose of this study is to use TTS and wave audio files to increase the accuracy of the existing language filtering. **Methods/Statistical Analysis:** In the Default setting, the text of the expletive BASE data that becomes the filtering standard is stored on the DB using TTS that reads the text into audio and the visualization program that produces the audio file into a wave file. When the text is inputted by the user, the device reads the text through the TTS Program, and the speech is visualized to be produced into a wave file. The inputted text is searched in the base data. The produced wave file and the wave file of the base data are compared, and when the match rate is 80% or higher, it is inserted into the DB. **Findings:** The proposed system is for increasing the accuracy of the language filtering of the expletives which are used in the online chat of online games. There were various studies on the expletives, but for the filtering on swear words, the swear filter is mostly used. However, there are limits in detecting swear words that are modified. Therefore, it is expected that this system will enable the improvement of detecting modified swear words to lead to healthy use of language through online. **Improvements/Applications:** It is considered that further studies are required to improve the expletive search engine to strengthen the restriction of expletive use, and to collect accurate and various expressions of expletive data to enable real-time, automated language filtering.

Keywords: Expletives, Language Filtering, Search Engine, TTS (Text-To-Speech), Wave File

1. Introduction

In the online used by many people, the use of indiscreet expletives is increasing more and more. Especially in online games where there are various age groups and mostly used by the youth, the use of expletives in the online game is becoming a serious social issue in terms of internet ethics. Also, verbal violence by misusing the anonymity of internet is getting more serious, and it is not only bringing chaos to the internet environment, but also inhibiting the use of appropriate language. However, the measures taken on this issue is insufficient, and it is only through the follow-up method after the report by the user, therefore, providing a system that can prevent the verbal violence in advance is a very urgent issue¹.

Currently there are many studies being conducted to solve this issue, and the subject most widely being studied is the language filtering system.

The language filtering system was developed along with the supply of internet to supplement the ethical issue of indiscreet uses of expletives through the internet. Regarding the studies related to language filter, many studies are on the expletive filtering system for preventing the use of expletives in user chatting on online games using SVM², on the system that converts the violent words displayed in the computer game and output on the screen³, and on the bulletin boards and chat program using filtering function⁴. There are many studies applied to the online games, and there are also studies on the swear word filtering system using the consonant and vowel alignment

*Author for correspondence

of the Korean alphabet to prevent the use of expletives even in general notice boards⁵. Along with the studies above, the commercialized language filtering searches the inputted text in the filtering DB, and when it matches the expletives that are stored, the inputted expletive text is block and outputted on the screen in substituted words to prevent the use of expletives. However, it cannot filter the expletives in modified expressions. Even though the developers are updating the language filtering data regularly, many expletives are still used through online. The users of these expletives cleverly change the consonants and vowels to create various forms of expletives, and there is a difficulty of having to add these modified words manually⁶. In other words, there are limitations in filtering all expletives including the modified expressions.

To supplement this problem, this study proposes the system of reinforcing the filtering function of various expletives by producing and comparing the wave files produced through the TTS (Text To Speech) that reads the letters. The overall flow of this process is as follows. First, the data that becomes the standard for the filtering is stored in the DB. The expletive texts that become the standard are inputted and the text is read through the TTS. Those audio files are stored in the image storage server in wave format, and the URL in which the text and the wave file are to be stored is stored in the DB. When the initial stage setting of the process is complete, the text inputted by the user is searched through the standard DB, and the data of the similar search result and the inputted text are compared with the TTS and with the wave file produced through the TTS. When the match rate is 80% or higher, it is determined as an expletive. The inputted text is stored in the DB in text and a wave file. When this process is performed, it is expected to improve the accuracy of the filtering on various expletives that cannot be filtered through the conventional language filtering system.

2. Language Filtering

Language filtering refers to the program that censors the expletives among the languages inputted by the users through the online cyber communication to inhibit the notification or indication through the game and internet to prevent any issue. Some online games are applied with the language filtering of outputting appropriate languages when the expletive words are used, and others are

applied to enable the language filtering of outputting by substituting the expletives with 'xx' or blank spaces. To this end, there were studies on a program that restricts the use of communication language and expletives operated through the client infrastructure and that recommends the use of appropriate expression⁷, considering the seriousness of cyber violence as a big issue.

In addition, language filtering is applied in many online games, portal sites and notice boards, such as in some internet notice boards, when there is an expletive word in the text, a message is shown on the screen to prevent the text from being registered⁸.

Generally, for the swear words used in the online game chat window, the filtering system of Swear Filter is being used⁹. Also in the online game portal site, the exclusive language filtering technique that applies this method is being used, but the performance is unsatisfactory, due to the vulnerability on modified swear words¹⁰.

2.1 TTS (Text-To-Speech)

TTS is a representative technology on voice synthesis, which is a technology that automatically produces the sound-wave of the voice. A voice of a person selected as a model is recorded and divided into certain voice unit, mark is attached to be inputted on the synthesizer, and only the required voice units are combined again to artificially create the voice. Because the text is converted to audio, it is also called the text-audio conversion system. In addition, there is a system that converts phonetic sign into audio¹¹. TTS is largely divided into front-end and back-end. The front-end performs two tasks, and first, when the number or elliptical expression is read inside the text, it is converted to expression. It is also called normalization of the text, pre-processing or tokenization. Second, each word is converted to phonetic sign to be divided into the metrical unit of text idiom, word or sentence, and sentence. The phonetic sign allocation work is called TTP (Text-To-Phoneme) conversion or GTP (Grapheme-To-Phoneme) conversion. In these two stages, the phonetic sign and the metrical information are combined to create a symbolized language expression, and it is sent to the front-end. On the back-end, the voice such as the meter is adjusted based on the result transmitted from the front-end to create more natural voice, and the actual voice data is outputted. In this process, the characteristics of the voice is fixed, therefore, the unique color of the voice synthesis software is shown. The field that mostly uses voice

synthesis is the screen reading software for those with difficulty in reading and for those who have difficulty in speaking¹² as the alternative method. The voice synthesis technology is also applied to the educational field. Studies are being conducted on the English learning program using TTS for the young and the general public who just started English¹³, and also on the foreign education of the voice synthesis such as the study on the voice synthesis of Sindhi and numbers¹⁴.

2.2 Praat

Praat is open software for computers on analyzing the scientific phonetic system of the voice. Praat means ‘voice’ in Dutch language. The software was developed by Paul Boersma and David Weeink of Amsterdam University, and it is being updated continuously. Also, the system can be executed in various OSs, and in a normal PC environment with a sound card, it is possible for use. Microphone can be used to record in mono or stereo sound, and the wave (wav) file used in IBM compatible model or Audio Interface File Format (aiff) used in Macintosh can be loaded for use. This program supports voice synthesis, and thus many users are using Praat in various fields regarding the voice analysis, music comparison, pronunciation and vocalization¹⁵.

3. Proposed Work

3.1 System Configuration Map

Figure 1 shows the overall configuration map of the system proposed in this study.

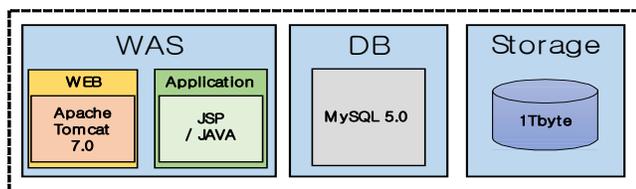


Figure 1. System configuration map.

The configuration of the proposed system is largely composed of WAS (Web Application Server), DB server and storage. For the WEB of WAS, Apache Tomcat 8.0 was used, and for the application, JSP/JAVA was used. For the DB server to store the expletive text and the URL of a wave file, MySQL Version 5.0 was used. Regarding the

storage that actually stores the wave file, the capacity was set to 1 TByte.

3.2 System Process

Figure 2 shows the overall configuration map of the system proposed in this study.

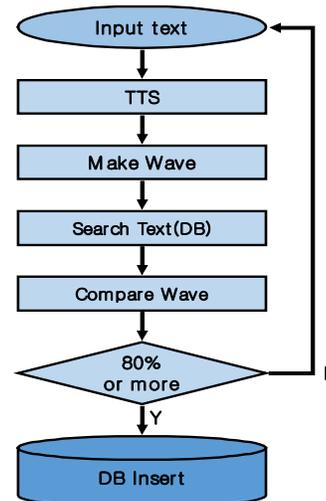


Figure 2. System process.

Default setting: It is the initial stage of the process. It is the stage of producing and storing the expletive text and a wave file that become the standard of filtering. The standard text is converted to voice, and produced in a wave file. Also, TTS tool is used to convert the text into voice. When converting to voice, the converted voice file is stored into the wave file. In the DB, the text and the URL of the relevant wave file are stored, and the wave file is stored in the storage server.

Step 1: It is the stage of the user inputting the text through the web, and the expletives and the sentences including the expletives are inputted.

Step 2: It is the stage of using the TTS tool to convert the text inputted by the user to voice. When the inputted text is converted to voice, the text inputted in the actor voice recorded in the TTS program is read through the voice. Through the tool, the voice is read always in the same condition. If a person directly reads the text, the data may be different depending on the person’s condition, but because the text is read in the same condition, the accuracy in the voice data can be expected.

Step 3: In this stage, the text is converted to voice to produce a wave file, and the file is stored in the storage. The TTS tool and the sound recording device are oper-

ated at the same time, and the text converted into the voice from the TTS tool is produced into the wave file by using the recording device. The produced wave file is used for comparing with the data that becomes the filtering standard of the expletives.

Step 4: In this stage, the text inputted by the user is searched through the DB, and the wave file is compared. The text inputted in Step 1 is searched in the DB stored with the Default information that becomes the filtering standard by using the Like sentence. The text of the data searched or the wave file, and the wave file produced in Step 3 are compared. The audio analysis program is executed for comparison, and the wave file searched from the DB and the wave file produced in Step 3 is loaded to compare the waveform.

Step 5: Using the audio analysis program in Step 4, when the match rate of the wave file is 80% or higher, it is determined as the new expletive to be stored in the DB. While the wave file is being stored in the storage, the overall flow of the proposed process is completed.

3.3 DB Table Configurations

Table 1 shows the DB table structure of the proposed process. The swear_words is the field name of the expletive text, and it becomes the value of the search and filtering standard to compare the expletive inputted by the user. The wave_img_url is the URL of the wave file on the text converted to voice and to be stored in the storage server. The wave file of the text converted from the actual voice is stored in the storage server. In the DB, the URL of the storage stored with the wave file is stored.

Table 1. DB table information

Field	Type	Description
idx_no	Int(11)	Index Number
swear_words	Varchar(50)	Swear Word
wave_img_url	Varchar(200)	Wave File URL
create_date	DATE TIME	Registration Date

4. Experiments and Discussion

In this study, to test the proposed system, the TTS tool provided by Oddcast was used in the stage of reading the text into audio. For the program on producing into

the wave file, the basic recorder program provided by Window was used, and for the tool to compare the wave files, the Praat program used as the audio analysis program was used.

4.1 Text-Audio Conversion using TTS

For the test, the TTS provided by Oddcast was used to input the text and to be converted to voice. When the program is executed, the conditions can be selected. The nation can be selected for the language selection, and the gender can also be selected. Because the country and the role of the voice actor can be selected in the preferred conditions, the preferred language can be selected to perform the expletive test. When performed after selecting the conditions, the inputted text can be converted to voice. In this study, the United States were selected for the country, English was selected for the language, and male was selected for the voice actor to perform the test. The standard of the text was set with the selected conditions, and various expletives were inputted for the test. Instead of the actual voice, the test was performed through the voice actor recorded without any changes and with the same voice to obtain the accuracy in the voice file. After inputting the selected conditions, the expletive text is inputted and converted to the voice. The converted voice is recorded to be produced and stored as a wave file.

4.2 Inputted Test Search in the DB

To determine whether the inputted expletive text is relevant to the expletives, it is searched in the DB using the Like sentence. The SQL searching the expletive text is shown in Figure 3.

```
SELECT swear_words, wave_img_url
FROM swear_word_filter_table
WHERE swear_words LIKE '%[input text]%'
```

Figure 3. Expletive data extraction SQL.

If there is a search result, the relevant wave file of the search result is loaded to be compared with the inputted text. If not, it is judged as not the expletives, and the process is completed.

4.3 Praat Wave File Comparison

To compare the wave file produced with the TTS program with the wave file in the DB, Praat was used as the audio analysis program in this study. The Praat program is used

to load the produced wave file and the wave file stored in the DB that becomes the filtering standard. Figure 4 shows the comparison of the wave file stored by using the Praat program and the produced wave file.

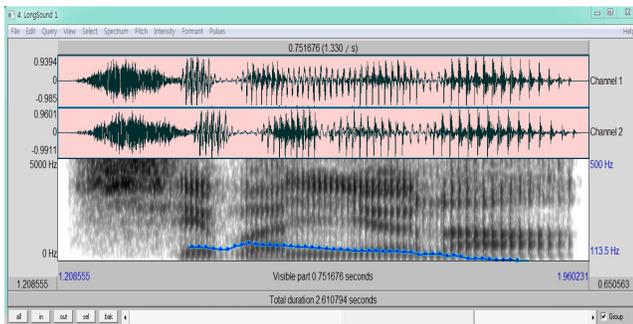


Figure 4. Wave file comparison.

The first wave is the waveform of the text that becomes the filtering standard loaded from the DB, and the second wave is the wave file of the inputted text for the test. The first and second waves show the similar wave forms, and the graphs of the register on the bottom almost coincide. The graph connected in the blue dotted line shows the graph of the register. It is judged whether it is an expletive or not by comparing the measures indicated on the left and the graphs shown on the bottom.

4.4 Data Input and Test

In the test order shown above, total of 500 expletives that were not modified were tested to obtain the test result as shown in Table 2. When the conventional filtering and the filtering proposed in this study were executed in the equivalent condition, the conventional filtering showed to be 97%, and the filtering of the proposed process showed to be 98%. The conventional filtering and the filtering of the proposed process all showed the high search success rate of unmodified expletives. Figure 5 shows the success rate of the unmodified expletive filtering.

Table 2. The filtering result of expletives that are not modified

Words Count	Before	After
100	95	96
200	193	195
300	291	296
400	387	392
500	489	490

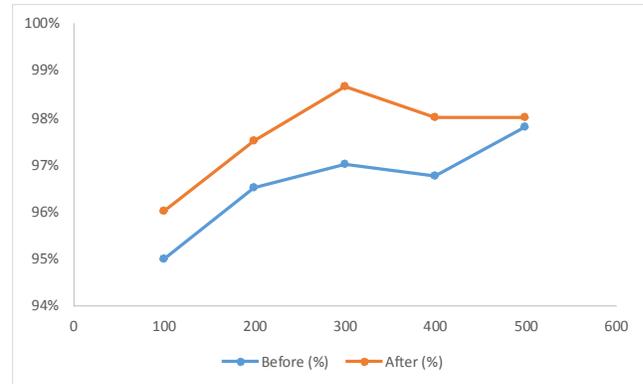


Figure 5. The filtering graph of the expletive expressions that are not modified.

Meanwhile, Table 3 shows the result values when the modified expression of expletive is inputted for comparison. The conventional filtering showed approximately 17% of success rate, but the filtering of the proposed process showed about 81% of success rate. In addition, 47% of improved expletive search success rate was verified. In the filtering of expletives in the modified expression, the filtering accuracy on the expletives was verified to be higher when the filtering in this study was used than the conventional filtering.

Table 3. Modified expletive filtering test result

Words Count	Before	After
100	19	82
200	38	161
300	50	247
400	64	312
500	77	419

Figure 6 shows the filtering success rate of the modified expletives. It was verified that the filtering accuracy was improved on the modified expletives in comparison with the conventional filtering.

5. Conclusion

In this study, to improve the conventional language filtering issue of not being able to filter the modified expletives, the process of reinforcing the search engine and improving the accuracy on the filtering was proposed.

It is expected that the process proposed in this study detects the expletives in a wider range than the conventional filtering, and also captures the expletives of various

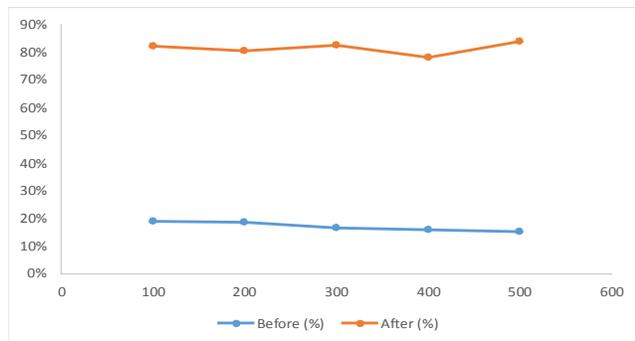


Figure 6. The filtering result of modified expletives.

modified expressions to help prevent the use of expletives. In addition, as languages per each country can be tested and inputted, the expletives of various languages can be filtered. As shown in the test and considerations, the filtering accuracy was improved than that of the conventional filtering, and the various expressions of expletives were verified to be filtered. In the future, the automated function should be added to the language filtering to automatically perform the proposed process per stage and be updated, and the collected data should be connected to the cloud service to share the data on the expletives globally to continue the study on preventing the use of expletives. In addition, it is expected that the language filtering is expanded not only in online games, but also to the portal site, notice board, mobile SNS and messenger to improve the issues of the internet ethics. In further studies, strengthening and increasing the accuracy of the expletive search engine for improving the restriction of expletive use, and also the study on the language filtering which collects various expressions of expletives data and is automated in real-time should be continued.

6. References

1. Yoon TJ, Cho HG. The online game coined profanity filtering system by using semi-global alignment. *The Journal of the Korea Contents Association*. 2009; 9(12):113-20.
2. Park KH, Lee JH. Developing a vulgarity filtering system for online games using SVM. *Proceedings of the Korean Institute of Information Scientists and Engineers Autumn*; 2006. p. 260-3.
3. Park KR. Study on establishment of internet ethics expressed through language convert in computer games. *Journal of the Korea Society of Computer and Information*. 2012; 17(11):47-52.
4. Koo KC. Planning and materializing information communication ethics experiencing study system, which filters the language of communication [Master's Thesis]. Korea: Busan National University of Education; 2002.
5. Yoon TJ, Cho HG. A filtering system for on-line vulgar words using Korean syllable alignment. *Proceeding of KIISE*; 2009; 36(2C):194-8.
6. Lee SWL. A swearword filters system for online game chatting. *Journal of the Korea Institute of Information and Communication Engineering*. 2011; 15(7):1531-6.
7. Jeon BK. The development application of agent recommending people to stop using slang expression of communication language [Master's Thesis]. Korea: Korea National University of Education; 2007.
8. Kim MG. Design and construction to intellectual slang expression filtering board system based on XML and .NET [Master's Thesis]. Korea: Sung Kyun Kwan University; 2003.
9. Swear Filter. Available from: http://en.wikipedia.org/wiki/Swear_filter
10. Shekhar D. Designing a vulgarity filtering system. *Game Programming Gems*. 2005; 5:621-6.
11. Cho JH, Kim TE, Lim JH. Implementation of TTS engine for natural voice. *Journal of Digital Contents Society*. 2003; 4(2):233-42.
12. Lim JH, Park EH, Lee SH. A preliminary study for korean text-based communication program development. *Journal of Special Education Research*. 2013; 12(1):247-73.
13. In J, Han J. The prosodic conditions in Robot's TTS for children as beginners in English learning. *Indian Journal of Science and Technology*. 2015; 8(S5):48-51.
14. Ratanpal BS, Sahni S. On speech synthesis of Sindhi numeric. *Indian Journal of Science and Technology*. 2015; 8(27).
15. Park HS. A visual study of the quality of english pronunciation using the Praat program. *Journal of Digital Contents Society*. 2013; 14(3):323-31.