

Explicit Words Filtering Mechanism on Web Browser for Kids

Nurul Athirah Binti Suliman
Faculty of Computing
Universiti Teknologi Malaysia
Johor Bahru, Malaysia
nurulathirahsuliman@gmail.com

Hazinah Binti Kutty Mammi
Faculty of Computing
Universiti Teknologi Malaysia
Johor Bahru, Malaysia
hazinah@utm.my

Abstract— Kids are introduced to and participated on internet usage from a very young age. Unfortunately, while they can build their skills to use technology quite speedily, this is not the case for their language prowess. There are double meaning words with one of its meaning being indecent or of an adult nature. As most kids cannot distinguish this, there are times when their innocent search leads them to unsuitable sites for their age. Furthermore, as most word filters are not equipped to efficiently filter these words, this problem is compounded. We propose a browser extension that can effectively and accurately filter profanity, bad words and words with double meaning. This solution can better provide a safer environment for kids to traverse through the Internet.

Keywords—component; bad words; kids; extension.

I. INTRODUCTION

In this day and age, people are constantly attached to the Internet for many different services, with information searching being the most popular. Since the technology has been growing rapidly; added with the multitude of different content that can be found ranging from presentations to videos, even short documentaries and films; this fact is inarguable. The Internet of Everything (IoE) is one of the example that shows people constantly connected to data, process and things to make network connections more relevant and reliable [1]. People can communicate with each other, search for information, send and receive data and also participate in social network.

Kids are introduced to the Internet at an earlier age today and their participation in it is even encouraged in schools or the purpose of learning. Kids, being young and innocent tend to search for everything they want in the Internet without really weighing on the good or bad of the search results. A weakness of internet browsers is appropriately-aged content filtering. Search engines does its job well from the point of view of searching but it is not very well equipped in content filtering, especially those age appropriate search results display. This shortcoming is glaringly worrisome when many sites have a very different content displayed to what seem to be an innocent enough site name or title. There are many words in the English language that have double meanings, for example crane, date, naughty, pet and toy. These words are somewhat harmless enough words which are used by kids, but regrettably some of them have a more 'adult' meaning to it too. Kids searching with these words, a variant of it or a combination of it, may be directed to inappropriate or adult sites. Most web filtering

software is not garnered specifically for different aged kids and the possibility of an innocent word leading to inappropriate content. Hence, an efficient and accurate web filtering approach is needed in order to filter explicit words for kids during internet search.

II. INTERNET SEARCH

In these times of technological advancement, internet can be accessed easier using computers, smartphones or tablets. Internet has become more important in our life as the main platform to search information. Internet will undergo many processes to give the information requested by user. There are few steps needed to search information through web browsers. First, select a search engine such as Google, Yahoo or Bing. Then, type a keyword(s) or information needed in the search box provided by the chosen search engine. Most search engine will even provide suggestions to help you get the best results. When you have entered your choice keywords, a list of web pages related to the keyword(s) requested are displayed and user may access the results to find accurate information.

You can be as generic or specific as you want in the keyword(s) that you type. The more specific the input criteria, the lesser number of results accumulated; the better to find exactly what you are looking for. If you type "internet safety" into the search bar (we are using Google), you will be given 177 million possible sites as a result. If you tweak this to be "internet safety for kids", you will get a result of 8.9 million. Tweak the search further to "internet safety for preschool kids", and you will get about 3.8 million results.

Some browsers, like Google, also group certain searches into specific tabs such as images, videos and news. We conducted a little experiment to find out the results received from these different tabs in a search, all within the current monitor display (which is 14.5") without scrolling down. We typed the phrase "naughty kitty" in a Google search engine. Table 1 shows the different outputs under each tab both when Safe Search is turned ON and OFF. It is interesting to note that while the first few search results under the 'all' and 'videos' tabs were quite harmless, this is not the case with the 'images' tab. An innocent kid may be subjected to age-inappropriate images, which if clicked on either intentionally or not, may bring them to an inappropriate site.

Table 1: Different search results under different search tabs

Keyword searched: naughty kitty Monitor size: 14.5"		
Tab	Safe Search ON	Safe Search OFF
All	5 links all age-appropriate	5 links all age-appropriate
Images	22 images : 4 age-inappropriate, 2 which can go either way, and 16 age-appropriate	22 images : 8 age-inappropriate, 2 which can go either way, and 12 age-appropriate
Videos	5 links all age-appropriate	5 links all age-appropriate

III. FILTERING

According to Whitenton [2], filter is one tool that can analyze a given set of contents which can exclude items that do not meet certain or selected criteria. Filtering also may help users narrow the scope from a large set of search results, but only if the filter matches the selected criteria that have been set in the extension of the web browser. These days, filtering is most crucial during web surfing especially for teenagers and kids. Parental supervision is also important to ensure their kids stay safe in cyberspace. Additionally, web browsers such as Chrome, Firefox, and Internet Explorer also provide parental control to help parents instead of having to physical hover behind your children at all time. Security features implemented in these browsers cover many aspects such as option to add site exceptions, control cookies, safe browsing mode and block access to certain sites. Table 2 shows some of the existing security features in these web browsers.

Table 2: Security Features in Web Browser (adapted from [3])

Browser	Security Feature
Chrome	<ul style="list-style-type: none"> Turn off Chrome's ability to save password. This can prevent kids from surfing certain sites that has been shared by parents. Under content settings, content can be managed in terms of images, and cookies. Certain plugin also can be disabled to make certain sites unusable.
Firefox	<ul style="list-style-type: none"> Un-tick remember password for sites. This can prevent child from having a password-protected online hiding place. Firefox meshes well with the security that has been set-up in Windows. Stop unwanted downloads, website filtering settings and block certain users from certain sites.

Internet Explorer	<ul style="list-style-type: none"> Block certain websites by adding URL to the block list. Allow the content settings to be changed which include images and active content to run in computer.
-------------------	---

An online news and review magazine *techradar*, listed the best browser of 2017 with Google Chrome leading the pack [4]. Mozilla Firefox and Microsoft Internet Explorer were at places 4 and 6 respectively, in a list of 7 placers. It is acknowledged that Chrome and Mozilla web browser among the safer web browsers which provides more features for filtering. Other than that, Windows also provides security features by limiting content that can be viewed by children. According to Microsoft, parents can block or allow specific websites by signing in family on the Microsoft account website. In addition, Chrome also provide Safe Search for Kids to provide search filtering that is restricted, and prevent inappropriate images, text from being displayed. There are many filtering types available and often used in website which are web filtering, content filtering and also hybrid of web and content, web content filtering.

A. Web Filtering

Web Filtering is used to block undesired Web pages by using screening of Web requests and analysis of the contents of the received Web pages [5]. Web filtering is important in order to search for more accurate information and it has many major applications. One of the major applications of the web filtering is protection against inappropriate content. The Internet is an important source to get more information. However, the Internet is also a major host that provides pornographic, violent and other contents that are inappropriate for most Internet users. So, in these cases, web filtering is used to block access to pages which are against a policy defined in the web browser.

Another major application for web filtering is preventing misuse of the network [5]. The main aim of web filtering here is to prevent and reduce misuse of the network resources in an organization. For example, free internet access for employees in an organization is misused because network connection also can be used for chat applications, network games, downloading streaming video and audio content. For these reasons, by using Web filtering system, we can protect against improper content and prevent misuse of the network in an organization [6].

On the other hand, web filtering applied at the firewall has disadvantages. This approach will slow down the network due to all incoming traffic pass through the firewall. Additionally, if the end-to-end encryption is used, web filtering cannot be performed because the traffic is encrypted through the firewall [5].

B. Content Filtering

Content filtering allows the request but the response is inspected at the proxy server [7]. Then, the actual payload of the packet is checked to determine if the payload contains anything that meets certain criteria. Payload is data that being

carried out within a packet or other transmission unit [8]. Allow and deny decisions are done after the actual payload is checked. Content filtering gives many added advantages to users as it provides the ability to block certain viruses, e-mail attachments, advertisements, cookies, and also pop-ups.

The content of the pages can be examined without knowing the server’s origin and certain contents of the pages can be removed which are intentionally allowed by URL blocking. These two famous filtering techniques, URL blocking and content filtering complement to each other. They offer a more complete Internet access solution for users. Content filtering does not always ensure accurate results [9]. Sometimes content filtering will block a website that is not causing any issues in the workplace.

C. Filtering Issues

The Internet offers a life without boundaries (at last in cyberspace) as everyone can search information without any age bias. However, it is very important for us to know whether the information received is accurate, fake information and not the kind sought. Two main issues during filtering process to search certain information in internet are performance and accuracy in providing information. The Internet may, through unbiased perspective, display articles or results that may have bad impacts to our society especially for the younger generation. These articles or results may cover critical issues such as pornography, adult material non-necessarily pornographic in nature, politics and religion.

However, this research focuses on the accuracy and potentiality of the internet giving results other than the kind sought. This relates to the children who use the internet for their own innocent purpose such as homework, play games and puzzles. These children may come upon sex sites due to tactics employed by some such sites to prey upon unsuspecting users. One of the tactics is by using search terms commonly used by children to direct unsuspecting internet users toward adult sites [10].

Browser extensions are small programs or plugins that can enhance the functionality of the browser. These extensions can be for a multitude of services, usually are very specific and direct. There are extensions for blogging like ScribeFire, shortening long URLs with Goo.gl, blocking ads with Adblock and many more. There are also security service specific extensions too. Table 3 below shows a few existing browser extensions available for the Chrome browser, built for bad word or profanity filtering.

Table 3: Existing browser Extension for profanity filtering in Chrome

Browser Extension	Overview
Profanity Filter	Eliminates profanity, racial slurs, sexual language and other offensive content from the web. It has personalized settings with a white list and blacklist words, which will be tagged and automatically blocked. It is a community-driven project, which relies on volunteers to expand the functionality.
WebCensor	Blocks swear words, bad, strong, foul language, swearing, cursing, etc. Has a predefine list of bad words to filter, while user can add their own custom words. Bad words found are changed to ***.
Safe Words	Filter and censor profanity. Works on Facebook, Twitter and YouTube! Using a smart algorithm, which they claim can filter/censor out 99% of bad words quite efficiently. Bad words are changed with some random characters like %@#%.
Bad Words Filter	Detects and blocks more than 340 obscene words. Compatible with various browsers. Though the demo only works for a list of 3 words. Works from a word list. It is not free.

Based on Table 3, it is shown that most of the filtering extension in Chrome only filters bad words whereas double meaning words are still not being filtered accurately. The Firefox browser extensions do not cover many add-ons to filter bad words. A positive aspect that some of these extensions have is the ability to add custom word list. Unfortunately, this is not very helpful with common words that have double meanings. Hence, this research plans to focus on web filtering of search terms commonly used by children which have double meaning.

D. Double Meaning Words Used by Kids

There are many double meaning words existing in the world. Double meaning words also can be categorized as double entendre which means a word or phrase that has two interpretations. There are words with double meanings and both are plainly innocent, such ‘bark’. Then there are such words where one meaning is innocent enough and another has a more adult or indecent meaning, like ‘screw’. There are few double entendre words often used by kids without knowing the other indecent meaning of the words used. This action may lead them to the inappropriate websites which may consist of pornography image or video, and other inappropriate content. Table 4 shows a couple of examples of words often used by kids with the indecent definition.

Table 4: Examples of Double Entendre Words Used by Kids (adapted from [11])

Word	Definition
Naughty	1. Disobedient badly behaved 2. Rude or indecent, typically because related to sex
Pet	1. A domestic animal for companionship and it will be treated with care and affection. 2. Sexually stimulating caressing or stimulating touching.

Terms like naughty and pet will relate to numerous pornography sites [10]. Figures 1 and 2 below show re two examples if kids typing the word ‘naughty games’. While adults may understand the different meaning, a child often called ‘naughty boy’ may think it is normal. The following figures show the use of search engines by using one of the double entendre word which is naughty. Example situation just like following:

1. Kids type “naughty games” in search engine.

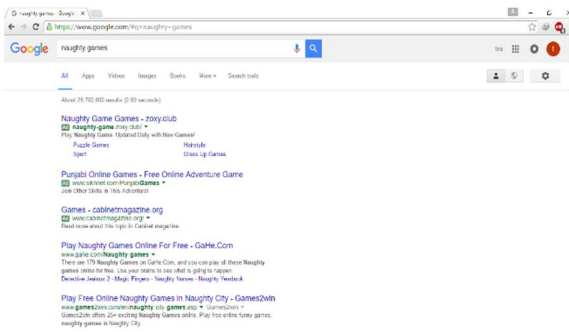


Figure 1: Search “Naughty Games”

2. The ‘naughty games’ result will be displayed inappropriate contents and not suitable for kids.

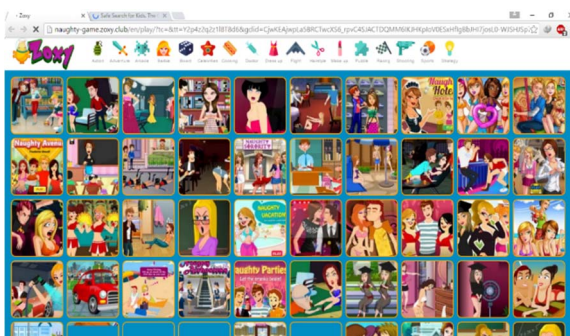


Figure 2: Naughty Games Result

IV. WEB FILTERING TECHNIQUES AND APPROACHES

Current implementations of Web filtering techniques are blacklisting or whitelisting, keyword searching and rating systems. Blacklist and whitelist are Web site list that must be allowed or blocked [5]. Blacklist will examine web sites manually and decide whether the website can be classified as a forbidden class member such as “Violence”. Sites also included in blacklists if their domain contains keywords such as “sex”. A list of allowed sites will be generated and other than that will be blocked. For blacklisting or whitelisting has an advantage of speed but also give some drawbacks which are costly to generate and hard to maintain.

Other than that, keyword filtering is used to block website based on words with keyword dictionary. If a page consists of a certain number of forbidden keywords, then the website cannot be accessed. This filtering technique is simple to implement but the disadvantages for this filtering is prone to spelling mistakes that can be used to bypass the protection [5]. Rating system such as PICS (Platform for Internet Content Selection) can generate rating for websites either in self-rating or third-party rating. Publishers of web page generate their own rating information for self-rating. While in third-party rating, independent third party is used to evaluate the web site and publish the results. In addition, rating system in web filtering generally does not provide a reliable and accurate source of information. Rating or labeling is the evaluation of the content of the website in terms of certain characteristics and factors such as sex and language [5].

Lastly, intelligent content analysis is one of the web filtering techniques which classify websites based on semantic understanding of context in terms of keyword they appear. The disadvantage of this filtering technique is that it is computer processing intensive. So, each technique has some disadvantages. Content analysis can be used by a web filtering system to classify automatically website through their contents [14].

A. Web Filtering Using Text Classification

Web Filtering by using text classification is widely used these days. This approach is used to determine the pages that must be blocked. Automatic text classification has been growing rapidly due to the increasing number of text documents in recent years [5]. Automatic text classification is one of a supervised learning task that assign predefined category labels to a new document by comparing it with a training set of labeled documents.

There are many approaches to text classification which are Naïve Bayes, K-Nearest Neighbour (KNN), Decision Tree (DTree), Support Vector Machines (SVM) and Neural Network [5]. Naïve Bayes (NB) are widely used because of this classifier is simple and provide computational efficiency. Relative frequency of words in the document will be used as word probabilities and these probabilities are used to assign a category to the document. Naïve Bayes also assumes that the probability

of words by a given category is independent for different value of words.

K-Nearest Neighbour (KNN) is one of statistical approach which is the most accurate methods of classifying many documents. Given a document, KNN will select most similar documents from the training set and use the categories of these documents to classify the categories of the document being classified. Documents are represented by vectors of words and the similarity between two documents which are measured using Euclidean distance. Another classifier which is Decision Tree is a machine learning approach and automatic induction of classification trees based on training data. In decision tree, each internal node is evaluated with a test on an attribute and outgoing branches of the node correspond to the results of the test. A part that associated with a category in decision tree is a leaf. Classification of a document starts from a root node and then visited the internal nodes until leaf is reached. The document category is the category of the final node.

In addition, Support Vector Machine (SVM) is also a classifier applied in text classification. Decision surfaces are used in SVM to divide data points into classes. Training documents represent as vectors and algorithm will determine hyperplanes to separate different classes of training documents. Test document will be classified according to their positions of the hyperplanes. Table 5 shows the differences between Naïve Bayes, Support Vector Machine and Decision Tree.

Table 5: Differences between Naïve Bayes, SVM and Decision Tree (adapted from [15])

Feature	Naïve Bayes	Support Vector Machine	Decision Tree
Speed	Very fast	Fast	Fast
Accuracy	Accurate in many domains	Accurate in many domains	Accurate in many domains
Transparency	No blackbox rules	No blackbox rules	Has blackbox rules

Naïve Bayes is among the fastest classifier in learning and predicting data. The accuracy of these three classifiers is accurate in many domains. Lastly, only decision tree has blackbox rules whereas the other classifiers have no blackbox rules.

B. Methods Chosen

Fewer items training data set are used, the nearest neighbour method [14] and Support Vector Machine will be more effective while said the Support Vector Machine is one of the best classification algorithm available [16]. So, among the many classifiers, Support Vector Machine is chosen to be applied in filtering explicit words.

Support vector machine uses a process to find a decision surface which can separate positive examples and negative examples in a multi-dimensional feature space [16]. Training documents in support vector machine are represented as vectors.

The following Figure 3 shows that hyper plane or known as support vector of SVM.

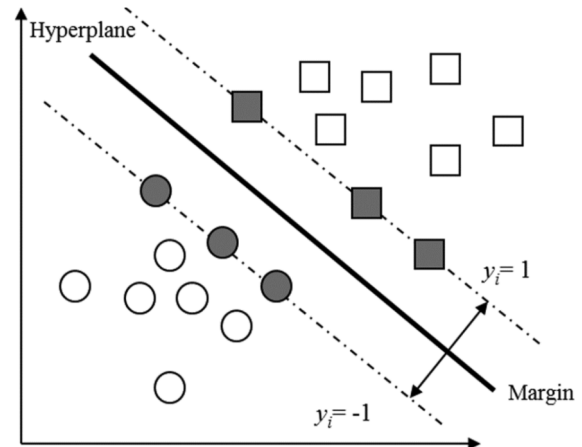


Figure 3: Hyper Plane of SVM (Chen & Hshieh, 2006)

The hyper plane which represented by bold line is separated into two categories and defined by the formula $(w.x) + b = 0$; where w is the weight vector and b is the bias that define the offset of the hyper plane from origin. The advantages of implemented SVM is error rate achieved is minimize since it uses a high dimension space to find a hyper plane to do binary division and it is efficient in handling a large number of training examples [17].

A. Data Collection & Feature Selection

Data Collection is the initial phase in Web filtering process. In this phase, data is collected by doing a survey and research about double entendre words that have been recently used by kids. After data has been collected, the data will be classified into different categories. The next phase will used this classified data for the next processes.

In feature selection phase, an appropriate term features are selected by using a term weighting schemes. In machine learning and statistics, feature selection or known as variable selection, feature reduction, attribute selection or variable subset selection is the technique of selecting a subset of relevant features [6]. It is used for building robust learning models.

Feature selection also based on reaction of the cross-validation data set classification error due to the removal of the individual features and feature selection is one of the special cases of feature extraction [18]. Every word in the text representation phase is treated as a term feature. Statistical weight approach is calculated using term weighting scheme to get the relationship of term feature and category of documents.

B. Term Weighting Scheme and Classification

Term Frequency-Inverse Document Frequency is a numerical statistic which shows how important a word to a document [6]. TFIDF is actually derives from TF and IDF which are very basic criteria in term weighting schemes. The TFIDF value will increase proportionally to the number of

times a word appears in the document. Term Frequency will measure terms that occurs frequently in document and Inverse Document Frequency will measure the importance of the term. For example, terms like “is” and “the” may appear a few times and not important terms. So, the terms that is used rarely will be focused on to measure the importance of a term in a document. TFIDF weighs the term or known as vector component of each documents which takes a few steps.

First step of TFIDF is calculating the term frequency in the document which give results on number of terms appear in the document and determine the more frequent terms appears, the term becomes more crucial. Inverse Document Frequency is used to measure the term whether it is common or rare in the documents. After that, the classification process will take place to classify the undesired web pages using Support Vector Machine (SVM).

An extension is a small software programs that can modify and the functionality is enhanced in the web browser. Extensions also has little user interface that can be implemented. Other than that, it will bundle all the files into one file so user can download and use this extension. Extension can implement to filter images, text or websites. In this research, an extension will be built to filter bad words recently used by kids.

V. CONCLUSION

This paper looked into and explained internet search and how words with double meaning can easily bring a child to inappropriate sites. We also considered a few different techniques that can be used for filtering bad words to find a suitable one to achieve the research objectives of always-appropriate search results for kids. The techniques are chosen based on the intention of enhancing the accuracy of data display based on users’ needs.

A. Limitation

Since certain web browser, such as Chrome does not allow the implementation of new extensions readily, we shall try to develop and apply our filtering extension to fit other viable browser.

REFERENCES

- [1] Dave Evans. 2012. The Internet of Everything on How More Relevant and Valuable Connections Will Change the World. Cisco, Amsterdam.
- [2] Kathryn Whitenton .2014. Filters vs. Facets: Definitions. University of Washington, Seattle.
- [3] Tiffany Carlson.2013. How to Enable Parental Control in Chrome, Firefox, Opera and Internet Explorer. Augsburg College, United States.
- [4] TechRadar. 2017. The Best Web Browser in 2017. <http://www.techradar.com/news/software/applications/best-browser-which-should-you-be-using-932466>.

- [5] Du, Safavi-Naini and Willy Susilo. 2003. Web Filtering Using Text Classification. The 11th IEEE International Conference on Networks,325-330.
- [6] Nurfazrina Binti Mohd Zamry.2013. Islamic Web Pages Filtering and Categorization. Master Thesis, Universiti Teknologi Malaysia.
- [7] Scott Jones .2002. URL Blocking vs. Content Filtering. Kansas State University.
- [8] Margeret Rouse.2005. Payload Definition. Nazareth College, United States.
- [9] David Romako. 2016. Some Advantage and Disadvantages of Content Filtering Rituals you should know in 2016. Arizona State University.
- [10] Crystal Roberts. 1999. Internet Filtering and Blocking Technology, The Most Effective Methods of Protecting Children from Pornography. Family Research Council Conference on Introducing the Children’s Internet Protection,America.
- [11] Oxford Dictionary. 2016. Retrieved 7th May, 2016 from <http://www.oxforddictionaries.com/>
- [12] Crystal Roberts. 1999. Internet Filtering and Blocking Technology, The Most Effective Methods of Protecting Children from Pornography. Family Research Council. Conference on Introducing the Children’s Internet Protection, America.
- [13] Lee, Z. S. 2010. Enhanced Feature Selection Method for Illicit Web Content Filtering. Ph.D. Thesis, Universiti Teknologi Malaysia
- [14] Yang, Y. and Liu, X. 1999. A Re-Examination of Text Categorization Methods, Proceedings of the 22nd International ACM SIGIR Conference on Research and Development in Information Retrieval. pp. 42–49
- [15] Anton Heijs. 2012.Support Vector Machines(SVM)-Text Analytics Algorithm Induction. Delft University of Technology, Netherlands.
- [16] Sebastian, F. 2002. Machine learning in automated text categorization. ACM Computing Surveys, Vol 34, 1–47.
- [17] Lin, P. C., Liu, M.D., Lin, Y.D., And Lai, Y. C. 2008. Accelerating Web Content Filtering by the Early Decision Algorithm. IEICE TRANS. INF. & SYST. Vol.E91–D, No.2, pp-251-257.
- [18] Verikas, A. and Bacauskiene M. 2002. Feature Selection with Neural Networks, Pattern Recognition Letters, 23,1323–1335.