Full Length Research Paper

Theoretical and experimental studies on textual anti-spam filtering

Mahathir A. Anwar[1], Lee Choo Bernard[1], Abdullah H. Victor[1] and Najib Jimmy Ibrahim[2]*

[1]Faculty of Engineering, Multimedia University, 63100 Cyberjaya, Selangor Darul Ehsan, Malaysia. [2]Department of Educational Management, Planning and Policy, Faculty of Education, University of Malaya, 50606 Kuala Lumpur, Malaysia.

Spam is unsolicited bulk messages sent indiscriminately. According to Wikipedia and Cisco report, more than 31 trillion spams have been sent in 2009. These spam or "junk mails" can involve various kinds of messages such as commercial advertising, pornography, viruses, doubtful product, get rich quick scheme or quasi legal services. In this paper, a direct attention has been paid to the text spam, and in particular, the process of text spam and the tricks of the spammers have been described in this paper. Moreover, the author described the implementation of the text content analysis and classification, using different document processing techniques (that is, stop words, short words form, regular expression, stemming etc.) and naive Bayesian classifier. In addition to that, the author has depicted the practical work of the document processing and naive Bayesian classifier towards implementing an accurate anti-spam system.

Key words: Text spam, stop words, short words form, regular expression, stemming, document processing, naive Bayesian classifier.

INTRODUCTION

The world economy is currently transitioning from a goods based economy to an economy of value creation, employment and economic wealth (Erbil and Akincitürk, 2010; Yass et al., 2010). As the amount of products and services offered via the internet grows rapidly, consumers are more and more concerned about security and privacy issues (Abomhara et al., 2010a, b; Zaidan et al., 2010b, c, d, e, f; Hashim, 2010, Alam et al., 2010; Naji et al., 2009). Nowadays, the internet is not only a networking media, but also as a means of transaction for consumers at the global market (Delafrooz et al., 2009). Potluri (2008) said "creating effective communication with customers is the most important aspect in service market".

Within global markets, the issues that businesses encounter are getting more and more complicated and sophisticated (Hsu et al., 2010; Alam, 2009b). High speed internet backbone has become ever-present and the high speed modems have become the standard entry-level in internet connection (Hmood et al., 2010b; Hmood et al., 2010b; Alam, 2009). Therefore, protecting the privacy of the data has become an urgent need (Al-Frajat, 2010; Zaidan et al., 2010a; Hmood et al., 2010c; Ahmed et al., 2010).

One of the well-known threats on network security is spam. Spam, social e-mails or junk mails are floods of e-mails sent to other e-mail boxes. Spam may contain many types of messages such as commercial advertising, doubtful product and pornography, get rich quick scheme or virus. Spam is mostly classified into direct mail messages or Usenet spam. Usenet spam is defined as a single message that is sent to twenty or more Usenet newsgroups. The Usenet targets the people in newsgroups, but rarely or never give their e-mail addresses away (Mueller, 1999). The other category is the direct

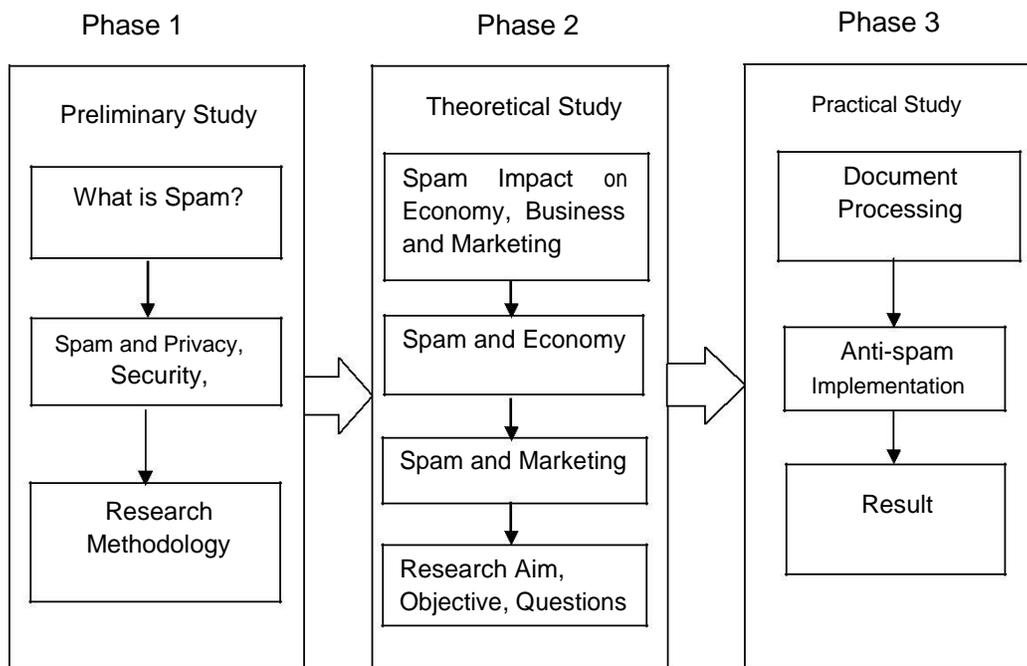*Corresponding author. E-mail: najib220@gmail.com

Phase 1   Phase 2   Phase 3

Figure 1. Research operational framework.

mail message which targets single e-mail messages by searching the web for addresses, stealing their mailing list and scanning the Usenet postings. The situation will get worse if the recipient reply to the spam messages and that will cause the recipients' addresses that will be available to be attacked by other spammers (Raad et al., 2010). Many researchers consider spam filtering as a type of text classification task. Spam filtering techniques have two categories which are general and specific (Westbrook, 2000). General filtering technique usually looks for suspicious elements in an e-mail such as capitalized subject field and obvious phony names "From" the sender or field including numbers and other features or exclamation points in the subject field, while specific filtering technique looks at the characteristics of actual spam messages (training processing), such as source of the spam, phrases or words, or specific requests from the recipient.

A popular example for the specific filter is the Bayesian filtering approach. Therefore, specific filtering technique appeared to be more accurate than general filtering technique since it requires more maintenance by the users, especially from the e-mail user(s) themselves. They depend on someone to set or classify the actual spam message to generate and update the spam signatures. To measure spam filtering, there are two parameters (effectiveness and accuracy used systems) (Westbrook, 2000). Effectiveness or true positive is measured by the percentage of spam that is wedged. The effectiveness percentage should be as high as pos-sible, while the accuracy percentage is measured by the percentage of e-mails that are identified incorrectly as a

spam. The second percentage should be as low as possible. According to Kosmopoulos et al. (2007), a well-designed spam filter should be properly achieved on both percentages. They described briefly two e-mail spam filters that employed different forms of the naive Bayes classifier and focused on the text of the messages. The ultimate goal of this effort is to measure the value added by non-textual features and more elaborate classifiers.

Spam impact on economy, business and marketing

Bulk electronic mail or spam, has become a key danger to internet efficiency. According to many researchers, more than 80% of the internet global e-mail traffic consist of spam messages. This increase causes bad storage capacity, like the unexpected overload of e-mail systems in bandwidth and loss of end-user efficiency. Therefore, spam messages had clear impacts on the economy, business and marketing. According to a report by MacAfee, entitled "MacAfee Americans and spam sur-vey", spam is the main technology time waster with 49% as compared to other technology-related annoyances including automated voice response systems (24%) and slow internet connections (19%). This survey exposed that 49% of Americans spend at least 40 min per week deleting spam, while 14% reported that they spend more than 3.5 h weekly deleting spam (Park et al., 2007). It is widely acknowledged that spam costs businesses large amount of money in terms of workforce productivity (Uys, 2009).

Spam and economy

Creating money from unsolicited e-mail is the main goal of spammers, either from other parties, such as porn or gambling sites or from their own scams or products (Sophos, 2005). From the economics perspective, European companies loss $2.8 billion in productivity because of spam, while US based companies reported a loss of $20 billion (Hinde, 2003; Raad et al., 2010). This loss includes the time of deleting the spam messages, the cost of increasing the storage of e-mail systems to handle the inboxes flooded with spam messages, and the networks overloaded by spam. Hinde (2002) states that junk e-mail has become a potent weapon used in targeting unsuspecting consumers, stealing their identities and money, and using their personal contacts (Hinde, 2002). Regardless of this, spamming can be economically viable because the operating costs for online communication channels are close to zero. The profit depends on the product, timing of the campaign, opening rate and purchase probability, which is influenced by the mail spam. Potentially, high profits and low market entry barriers continuously attract spammers; however, legal actions against spam have been started by legislation (Zhang, 2005).

Spam and marketing

E-mail marketing is one of the cheapest ways of advertising products. Since e-mail users are growing rapidly, more and more businesses are choosing e-mail marketing for their advertisement campaign. Recently, e-mail marketing is considered as a great method of reaching the global audience for their target market (Raad et al., 2010). In the new global investments and business revaluation, e-mail marketing service is playing an important role on advertising the products; however, a number of challenges face the usage of e-mail marketing service such as junk e-mail "spam" (Raad et al., 2010; Uys, 2009).

Research aim and objectives

In the literature, spam problems and its influence have been investigated and discussed from different perspectives. Several researchers have looked into the influence of spam on economy, finance, marketing, business and management, while other researchers studied the impact of spam on security, privacy and date protections. Moreover, there were many researches that spotted a light on anti-spam filter techniques such as machine learning and IP blocks. In addition to that, there were several researches that were conducted to illustrate the impact of spam on the society, spam and law, and spam and e-mail reliability.

There are two main objectives in this research. First, the theoretical part where the impact of spam on economy, finance, marketing and business will be studied; secondly, the practical part where multi-functions of document processing such as word stemming, short message form, stop words and tokenizing will be created. Moreover, an adapted machine learning (Bayesian method) working together with the study's document processing to implement an accurate anti-spam engine will be used.

RESEARCH QUESTIONS

1. What are the stop words, and how important are the stop words in reducing time cost and increasing accuracy of the spam filter?
2. What are "short form words", and can these words affect the accuracy of the classifier?
3. What is word stemming?
4. How accurate is naive Bayesian classifier?
5. Is there any standard for the stop words, regular expression, short form words, stop words and word stemming?

Spam system implementation

The anti-spam system that implemented the use of Machine Learning (ML) consists of three stages which are: pre-processing, training and testing (Figure 2).

Document processing

Several contents of the e-mails have to be processed before it can be applied in the appropriate algorithm. Therefore, to apply the algorithm in e-mail filtering and classification, first, contents of an e-mail should be transformed into numeric data. The main content of the data has a subject and body. Typically, the data in the header and the body of the e-mail is similar, although the header and the body can be dissimilar and do not point out exactly what the sender wants to inform the receiver. Therefore, using only headers can slightly reduce the accuracy of e-mail spam filter (Kiritchenko et al., 2002). This process is called preprocessing and it involves the process of feature extraction, reading and tokenizing, feature selections, stop word removal and stemming (Figure 3).

Reading and tokenizing

Tokenization is defined as the identification of the "atomic" unit, which represents the very first procedure in document processing; however, it is often overlooked because of its consideration as a basic nature. Even though
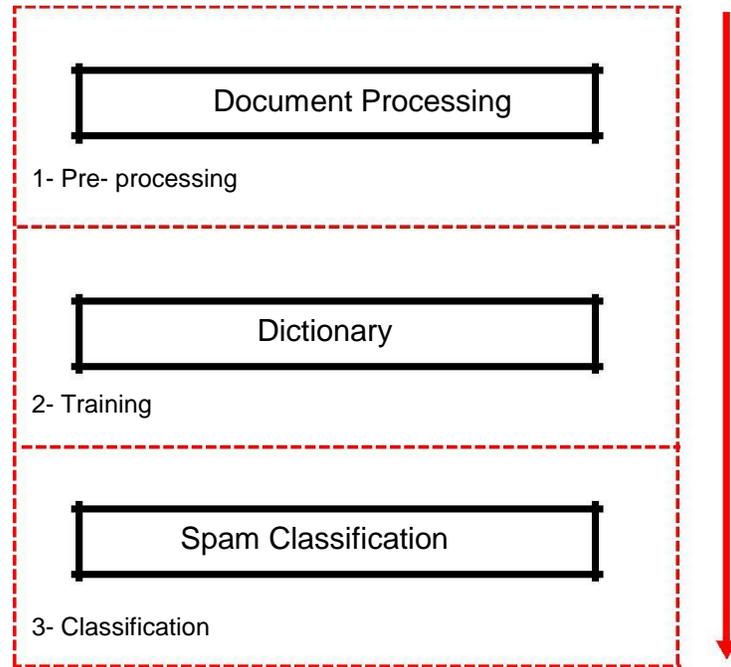
Document Processing

1- Pre- processing

Dictionary

2- Training

Spam Classification

3- Classification

Figure 2. Spam detection system.

Start Document Processing

Short Words Form

Stop Words

Regular Expression

Tokenizing

Features Counting

Word Steaming

Content Scanning

Sub Words

Words

Domain html

Features

Sub Features

Black list

Domain-frequency

Domain html frequency

Domain Subject frequency

Feature frequency
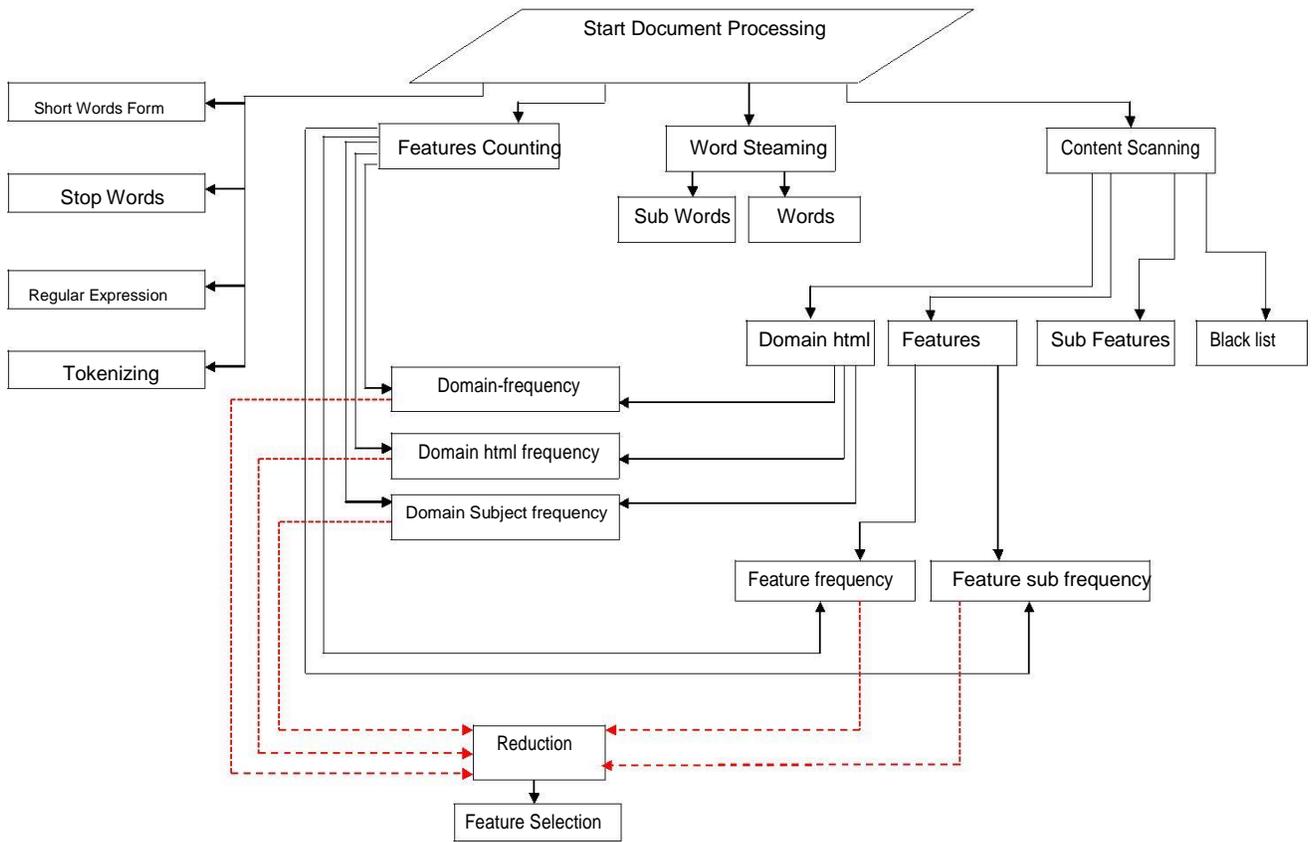
Feature sub frequency

Reduction

Feature Selection

Figure 3. Depiction of the most helpful document processing before the classification.

Table 1. Some of the Arabic examples used in the chat.

Character in Arabic	Character in English	Character in Arabic	Character in English
ئ,ء,أ	2	ب	b
ت	t	ث	th
ج	g	ح	7
خ	5	د	d
ذ	4	ر	r
ز	z	س	s
ش	sh	ص	9
ظ	'9	ط	6
ظ	'6	ع	3
غ	'3	ف	f
ق	8	ك	k
ل	l	م	m
ن	n		h
و	w	ي	e

Table 2. Three sentences in English, Arabic and modern Arabic.

English	Arabic	Modern Arabic chat
The student went home"	▓▓▓▓▓▓▓▓	"4ahaba al6aleb 2ela albeit"

the obvious simplicity of the issue is at stake, there is no available standard or solution that exists for the character stream tokenization. There is no general agreement on even a meager definition of this stage, and this leads to the lack of shared techniques and knowledge in this area. Moreover, very little attention is paid to the appraisal of the quality of the result. So far, no evaluation methods (that is, metrics) have been designed and used in this particular area of NLP (that is, natural language processing) (Habert et al., 1998). Practically, e-mails in text format should be transformed into a vector before the tokenizing process acts on these e-mails. In the tokenizing process, all the symbols such as (},.:, % $ @) will be deleted, and only the words which contain characters A to Z or a to z, will be saved in this vector or in the new vector. This process can make the classification process easier. The output of the tokenizing process is unique in words only or in terms of a row. Another problem on this area appeared when the spammer uses another language, such as Arabic spam. The modern Arabic messages used English characters with numbers. These numbers were used instead of some Arabic letters which did not appear in English as shown in Table 1. An example for the modern Arabic chatting language can be shown in Table 2.

This technique has become a popular chat language. In addition to that, spammers started using it. According to the study's understanding, there is no single paper that talked about this kind of spam; however, one e-mail that used this technique has been caught at the author's e-mail.

Finally, the same technique can be used in English language or in any other languages. Particularly, short words message in the English chatting language is a good example for the future of the spammer's techniques.

Short words form

Unlike SMS (short message service), where the user uses the short form due to the limitation of the message size, spammers start using the short forms to confuse the spam engines. According to Fung (2005), a few companies create SMS translation software, and one of these companies is Geneva Software Technologies Limited (GSTL). This new SMS traditions have initiated several websites such as Canada's transl8it.com, which offered SMS translation services through direct word to word matching. Moreover, cooperation was encouraged among service providers and translation companies such Singapore's GistXL Pte Ltd to come up with software such as GistXL, and Simplified Chinese SMS translation platform embedded in SingTel's Singapore network. As far as we know, researches are yet to be done on translating the short words form into long forms for the purpose of overcoming the problem of confusing the anti-spam engine. As seen in the aforementioned, stop words and short form messages need to have a long study. In the stage of document processing, the anti-spam will either remove the unknown word or consider this word as

Table 3. Examples of short messages forms.

?	I have a question	4	Short for "for" in SMS
?	I do not understand what you mean	411	Meaning "information"
?4U	I have a question for you	404	I do not know
;S	Gentle warning, like "Hmm? What did you say?"	411	Meaning 'information'
^^	Meaning "read line" or "message above"	420	Lets get high
<3	Meaning "sideways heart" (love, friendship)	420	Meaning "Marijuana"
</3	Meaning "broken heart"	459	Means I love you (ILY is 459 using keypad numbers)
<33	Meaning "heart or love" (more 3s is a bigger heart)	4COL	For crying out loud
@TEOTD	At the end of the day	4EAE	Forever and ever
0.02	My (or your) two cents worth	4NR	Foreigner
121	One-to-one (private chat initiation)	^5	High-five
1337	Leet, meaning 'elite'	511	Too much information (more than 411)
143	I love you	555	Sobbing and crying. (Mandarin Chinese txt msgs)
14AA41	One for all, and all for one	55555	Crying your eyes out (Mandarin Chinese txt msgs)
19	Zero hand (online gaming)	55555	Meaning laughing (In Thai language the number 5 is pronounced 'ha'.)
10X	Thanks	6Y	Sexy
1CE	Once	7K	Sick
1DR	I wonder	831	I love you (8 letters, 3 words, 1 meaning)
2	Meaning "to" in SMS	86	Over
20	Meaning "location"	88	Bye-bye (Mandarin Chinese txt msgs)
2EZ	Too easy	88	Hugs and kisses
2G2BT	Too good to be true	9	Parent is watching
2M2H	Too much too handle	<s>	Meaning "smile"
2MI	Too much information	*s*	Meaning "smile"
2MOR	Tomorrow	*w*	Meaning "wink"
2NTE	Tonight		

unknown. For instance, the word ROFL, is meaningless for the spam engine, while it actually means "Rolling on the floor laughing". According to Beal (2010), they publish more than 1300 abbreviations. As can be noticed from Table 3, the short form of the word is a new challenge to the anti-spam. A table of 1300 short forms can be found at Beal (2010) in http://www.webopedia.com.

Feature extraction

The first step in document classification process is feature extraction. This task (that is, features extraction) has an important influence on the learning process. Thus, the sentences from the training document will be extracted. The main goal of doing the feature extraction is to decrease the dimension column from management information problem. On the other hand, the extraction process could also improve the accuracy of the classification execution. However, this process has an

impressive effect on the unstructured text because of its potential text column 'string' type, which is infinite and as such, the word meaning changed following a particular word that it contains. Therfore, the technique of reducing the number of features has been applied to decrease the total dimension of the list number which was done earlier. The result of this list is called "The Dictionary". The most popular methods that are usually used to reduce the dimension list number are stop word removal and word stemming.

Stop word removal

The first step in the extraction of the feature process is stop word removal. This function works by detecting the stop word; however, stop words consist of unused words that are included in the text, which appear in approximately 5% of the documents. According to several researches, there is a short list of common words such as:

Table 4. Sample of stop words.

a	did	herself	not	the	we've
about	didn't	him	of	their	were
above	do	himself	off	theirs	weren't
after	does	his	on	them	what
again	doesn't	how	once	themselves	what's
against	doing	how's	only	then	when
all	don't	i	or	there	when's
am	down	i'd	other	there's	where
an	during	i'll	ought	these	where's
and	each	i'm	our	they	which
any	few	i've	ours	they'd	while
are	for	if	ourselves	they'll	who
aren't	from	in	out	they're	who's
as	further	into	over	they've	whom
at	had	is	own	this	why
be	hadn't	isn't	same	those	why's
because	has	it	shan't	through	with
been	hasn't	it's	she	to	won't
before	have	its	she'd	too	would
being	haven't	itself	she'll	under	wouldn't
below	having	let's	she's	until	you
between	he	me	should	up	you'd
both	he'd	more	shouldn't	very	you'll
but	he'll	most	so	was	you're
by	he's	mustn't	some	wasn't	you've
can't	her	my	such	we	your
cannot	here	myself	than	we'd	yours
could	here's	no	that	we'll	yourself
couldn't	hers	nor	that's	we're	yourselves

1. Articles such as "a, an, the"
2. Prepositions such as "or, and"
3. Number and dates
4. Word that shows connectivity (and, but, because).

Removing these words will save spaces in order to store the contents of the document and reduce the search process time (Selamat et al., 2003). As was previously mentioned, stop words are based on word frequency. For instance, here are some common words which are included in the stop word list (Table 4). According to many researches, there are other tens of stop words which can also be excluded at this stage. The study's system is flexible in adding and removing more stop words. Thus, it is recommended that the stop words and the short words form should be studied linguistically.

Word stemming

Word stemming is defined as a common form of natural language processing in most of the information retrieval (IR) systems. It is an important feature supported by the recent search systems and indexing. The idea of word stemming is to enhance the recall by automatic handling of word endings. This process can be done by reducing the words to their roots and is usually done at the time of searching and indexing. Stemming, typically, can be done by removing any attached suffixes and prefixes from the indexed terms and its process eventually amplify the retrieved documents number. In NLP, lumping or merging together non-identical words is called "conflation", and it is usually accepted that word-endings (that is, suffix stripping) removal is a good idea, while removal of prefixes can be valuable in some of the subject domains. However, it is not commonly practiced. The stemming process is the second feature extraction after the stopping process. It is done in sequence to eradicate the word-prefix and word-suffix, in order to get only the word root. Typically, the stemming process is applied only on English language at the suffix words such as "studying" and "helping", thus, some modification of stemming algorithm need to be done before the stemming process

spam techniques, internet control limitation and high speed bandwidth.

Conclusion

Spam is defined as a message sent from a user to another user without having to appreciate the receiving of this message. In this paper, naive Bayesian classifier has been described theoretically and practically. Moreover, the available document processing such as word stemming, stop words and short words form were studied. The study's anti-spam engine was successful in achieving around 96.36% with approximately 1.82% false positive and almost 1.82% false negative. Further study can still be done on analysing the importance of word stemming, stop words and short messages form on the accuracy of the anti-spam, and another research area can be created using anti-spam engine for different languages. Moreover, a multi language anti-spam filter might be created using the same document processing with the Bayesian method or any other machine learning. Another research that can be extended from this research is to study linguistically the stop words, short messages form and the stemming of other languages towards implementing content-bases text spam for other languages.

ACKNOWLEDGMENTS

REFERENCES

Abomhara M, Khalifa OO, Zakaria O, Zaidan AA, Zaidan BB, Alanazi HO (2010a). "Suitability of Using Symmetric Key to Secure Multimedia Data: An Overview." J. Appl. Sci., 10(15): 1656-1661.

Abomhara M, Khalifa OO, Zakaria O, Zaidan AA, Zaidan BB, Rame.A (2010b). "Video Compression Techniques: An Overview." J. Appl. Sci., 10(16): 1812-5654.

Ahmed MA, Kiah MLM, Zaidan BB, Zaidan AA (2010). "A Novel Embedding Method to Increase Capacity and Robustness of Low-bit Encoding Audio Steganography Technique Using Noise Gate Software Logic Algorithm.", J. Appl. Sci., 10(1): 59-64.

Alam GM, Kiah MLM, Zaidan BB, Zaidan AA, Alanazi HO (2010). Using the Features of Mosaic image and AES Cryptosystem to Implement an Extremely High Rate and High Secure Data Hidden: Analytical Study. Sci. Res. Essays, 5 (21): 3254–3260.

Alam GM, Khalifa MTB (2009). The impact of introducing a business marketing approach to education: a study on private HE in Bangladesh. Afr. J. Bus. Manage., 3(9): 463-474

Alam GM (2009b) Can governance and regulatory control ensure private higher education as business or public goods in Bangladesh? Afr. J. Bus. Manage., 3(12): 890-906

Alanazi HO, Alam GM, Zaidan BB, Zaidan AA (2010). "Securing Electronic Medical Records Transmissions over Unsecured Communications: An Overview for Better Medical Governance." J. Med. Plants Res., 4 (19): 2059–2074.

Al-Frajat AK, Jalab HA, Kasirun ZM, Zaidan AA Zaidan BB (2010). "Hiding Data in Video File: An Overview." J. Appl. Sci., 10(15): 1644-1649.

Beal V (2010) Text Messaging and Chat Abbreviations: A Guide to Understanding Text Messages, Chat Abbreviations, and Twitter Messages.

Erbil Y, Akincitürk N (2010). "An exploratory study of innovation diffusion in architecture firms." Sci. Res. Essays., 5(11): 1392-1401.

Fung LM (2005), "SMS Short Form Identification and Codec", Unpublished master's thesis, National University of Singapore, Singapore. 32 pp.

Habert B, Adda G, Adda-Decker M, Marëuil PBd, Ferrari S, Ferret O, Illouz G, Paroubek P (1998). Towards Tokenization Evaluation. In First International Conference on Language Resources and Evaluation (LREC'98). Grenade, Espagne, ELRA, 427-431.

Hashim F, Alam GM, Siraj S (2010). "Information and communication technology for participatory based decision-making-E-management for administrative efficiency in Higher Education." Int. J. Phys. Sci., 5(4): 383-392.

Hmood AK., Jalab HA., Kasirun ZM, Zaidan AA, Zaidan BB (2010a). "On the Capacity and Security of Steganography Approaches: An Overview." J. Appl. Sci., 10(16): 1825-1833

Hmood AK., Jalab HA., Kasirun ZM, Alam GM, Zaidan AA, Zaidan BB (2010b) "On the accuracy of hiding information metrics: Counterfeit protection for education and important certificates." Int. J. Phys. Sci., 5(7): 1054-1062.

Hmood AK, Zaidan BB, Zaidan AA, Jalab HA (2010c). "An overview on hiding information technique in images." J. Appl. Sci., 10(18): 2094-2100.

Hinde S (2002), "Spam, scams, chains, hoaxes and other junk mail." Comput. Security., 21(7): 592

Hinde S (2003), "Spam: The evolution of a nuisance." Comput. Security, 22(6): 474.

Kiritchenko K, Matwin S (2001). "E-Mail Classification with Co-Training". Proceedings of the conference of the Centre for Advanced Studies on Collaborative research (CASCON '01)

Kosmopoulos O, Paliouras G, Androutsopoulos I (2008). "Adaptive spam filtering using only naive bayes text classifiers", Appeared at CEAS2008 Fifth Conference on E-mail and AntiSpam.

Mueller SH (1999). "What is Spam", Referred on 16th October 2005 from World Wide Web: http://spam.abuse.net/overview

Naji AW, Zaidan AA, Zaidan BB (2009). Challenges of Hidden Data in the Unused Area Two within Executable Files. J. Comput. Sci.,5(11):890-897.

Park I, Sharman R, Rao HR, Upadhyaya S (2007). "The Effect of Spam and Privacy Concerns on E-mail Users' Behavior". J. Info. Syst. Security, 3 (1): 40-62.

Raad M, Yeassen NM, Alam GM, Zaidan BB, Zaidan AA (2010). "Impact of spam advertisement through e-mail: A study to assess the influence of the anti-spam on the e-mail marketing." Afr. J. Bus. Manage., 4(11): 2362-2367.

Sharma U, Kalita J, Das R (2002). "Root Word Stemming by Multiple Evidence from Corpus." In: International Conference on Natural Language Processing, pp. 31–39.

Sophos (2005) "The spam economy: the convergent spam and virus threats" whitepaper. May 2005 available at: http://www.sophos.com/whitepapers/Sophos_spam-economy_wpus.pdf

Uys L (2009). "Voice over internet protocol (VoIP) as a communications tool in South African business." Afr. J. Bus. Manage., 3(3): 089-094.

Westbrook B (2000). "The Basic of Spam Filtering". Referred on 17th of October 2005. URL: http://www.mail-filters.com

Yass AA, Yasin NM, Alam GM, Zaidan BB, Zaidan AA (2010). "SSME Architecture Design in Reserving Parking Problems in Malaysia." Afr. J. Bus. Manage., URL: http://www.academicjournals.org/AJBM (In press).

Zaidan AA, Zaidan BB, Al-Fraja AK, Jalab HA (2010a). Investigate the Capability of Applying Hidden Data in Text File: An Overview." J. Appl. Sci., 10(17): 1916-1922.

Zaidan AA, Zaidan BB, Al-Frajat AK, Jalab HA (2010b). An overview: Theoretical and mathematical perspectives for advance encryption standard/rijndael. J. Appl. Sci., 10(18): 2161-2167.

Zaidan AA, Zaidan BB, Alanazi HO, Gani A, Zakaria O, Alam GM (2010c). "Novel approach for high (secure and rate) data hidden within triplex space for executable file." Sci. Res. Essays., 5(15):1965–1977.

Zaidan AA, Karim HA, Ahmed NN, Alam GM, Zaidan BB (2010d). A New Hybrid Module for Skin Detector Using Fuzzy Inference System Structure and Explicit Rules. Int. J. Phys. Sci., 5(13). (In press).

Zaidan AA, Ahmed NN, Karim HA, Alam GM, Zaidan BB (2010e). Increase Reliability for Skin Detector Using Backprobgation Neural Network and Heuristic Rules Based on YCbCr. Sci. Res. Essays., 5(19): 2931–2946.

Zaidan AA, Karim HA, Ahmed NN, Alam GM, Zaidan BB (2010f). A Novel Hybrid Module of Skin Detector Using Grouping Histogram Technique for Bayesian Method and Segment Skin Adjacent-Nested Technique for Neural Network . Int. J. Phys. Sci., 5(14). (In press).

Zaidan AA, Zaidan BB, Taqa AY, Mustafa KMS, Alam GM, Jalab HA (2010d)" Novel Multi-Cover Steganography Using Remote Sensing Image and General Recursion Neural Cryptosystem." Int. J. Phys. Sci., 5 (11): 1776-1786.

Zaidan BB, Zaidan AA, Al-Frajat AK, Jalab HA (2010a). "On the Differences between Hiding Information and Cryptography Techniques: An Overview." J. Appl. Sci., 10(15): 1650-1655.

Zaidan BB, Zaidan AA, Taqa A, Alam GM, Kiah MLM, Jalab HA (2010b), "StegoMos: A Secure Novel Approach of High Rate Data Hidden Using Mosaic Image and ANN-BMP Cryptosystem." Int. J. Phys. Sci., 5(11):1796-1806.

Zhang L (2005). The CAN-SPAM act: An insufficient response to the growing spam problem. Ber.Technol. Law J., 20: 301-332.