Development of Scammed Posts Detector: A Case Study of Pet Scammed Posting

Nor Sa'datul Aqma NORAZMAN, Norshuhani ZAMIN
Department of Computer and Information Sciences
Universiti Teknologi PETRONAS
Bandar Seri Iskandar, 31560 Tronoh, Perak, Malaysia

ABSTRACT

This paper presents a research on scammed posts detector which focused on pet scammed posting detection. The research was motivated by the unawareness of pet lovers and owners on the pet scamming activities over the internet. Additionally, the current scam detectors are not able to notify the user on the potential pet scam posts they are dealing with. The objective of the research is to develop an application to automatically identify and alert the user on the potential pet scammed posting as an extension to dear current email system. The scope of the research are online pet advertisements and email communications between victims and scammers. Preliminary investigation on real pet scammed emails, experiences of the real pet scam victims and the public awareness regarding pet scams have been conducted to support the research development. Pet details, hardship of the family, obstacles faced and victim details are among the patterns extracted in the collection of real pet scammed emails. The system is tested against the human baseline and reached up to 86% of accuracy.

Keywords: Pet Scammed, Text Mining, Artificial Intelligence.

1. INTRODUCTION

Pet scams are one of the popular internet scams today. Pet scammed posting is a fake posting about pet sales or adoptions in the online auction sites. The pet sales or adoptions normally involve rare and exotic breeds such as Bengal cats, Beagle dogs and Macaw parrots that do not exist. The scam advertisements are posted in real websites and the dealing between the scammer and potential victim is done through the emails. Email communications are used to prevent the website administrator from tracing their illegal activities. The victims of pet scams are mostly the exotic pet lovers who are looking for adorable pets at a lower market price or pets which are looking for good home. The number of pet scams victims aroused due to the lack of knowledge and awareness on the scamming activities. They tend to commit to offers that are difficult to resists and sounds too good to be true. The Internet Crime Complaint Center (IC3) in the US has firmly established its role as a valuable resource for both victims of Internet crime and law enforcement agencies investigating and prosecuting cyber crime recently reported that they processed 262,813 complaints in 2013, representing more than $781 million in losses mostly in the classified advertisements[1]

Our literature found that the scammers claimed to give away the pets because of a family hardship such as relocation, death of pet owners, change of climate, etc [13,14]. Example of questions posted by pet scammers is shown in Figure 1.

Are you married?,,,,,,,,,,,,,,,,
Do you have kids?,,,,,,,,,,,,,,
Have you ever kept a puppy?,,,,,
Where are you located?,,,,,,,,,
What do you do for a living?,,,,
How soon do you want these puppies?,

Figure 1: Example of Questions in Pet Scams

As far as our knowledge is concerned, there is no automated online scammed posts detector which tailored to online sales in existence. Current scams detector like the Scam Detector App [1] provides only a guideline on how to avoid scammed posts with real examples but is unable to notify users when they might be involved in potential scammed deals. The main effect of pet scammed deals is a huge monetary loss as reported by the victims [2]. Money lost in internet crimes hits a new high last year, topping about $240 million, according to a government report showing increases in scams including pet scams [11]. As in the pet scammed cases, the scammers whom claimed to be resided in the other side of the world than the victims, offer some pets at very low prices (some are even free for adoptions) but requested the victims to pay for the shipment cost of the pets. The target victims are those countries in the world which has been statistically reported to have a high number of pet lovers such as in the United Kingdom [3], United State [4], Australia [5] and New Zealand [6].

Hence, the objective of this research is to develop an automated scammed posts detector which studies on pet scammed related cases. The scope of the research is to detect potential scams via the email communications between the scammers and the victims. The research takes a supervised learning approach which the categorization of the emails is based on the collection of real pet scammed emails. The outcome of the research is an extension in the email tool as an

[1] *http:// www.ic3.gov*

additional feature for email users to automatically removed pet scammed emails.

2. LITERATURE REVIEW

Statistically, more than half of the people in the UK keep a minimum of one pet and this has shown that they are pet lovers [3]. It is about 6.6 million dogs and 7.7 million cats is part of British household. In addition, in United State (US), 53.4 million U.S. households with more than one pet reported for 2012 [17]. There are about 95.6 million and 83.3 million of the number of owned cats and dogs respectively in US. In other countries such as Australia and New Zealand, 63% of the 7.5 million households own pets [18] and 68% of the households own at least one pet [19] respectively. Based on these figures, majority of the countries' household own at least a pet and there are exposed to the potential of pet scams risks. Meanwhile, the internet scams cases aroused from 100,000 reports per year to nearly 300,000 reports per year from 2005 to 2009. The monetary loss due to these is accumulated to exceed $300 million per year [10]. The statistic has shown that online users especially pet lovers have low scam awareness and ability to identify potential scams. This situation has motivated us to provide a solution to protect the online users.

Artificial Intelligence (AI) is a science that has defined its goal as making machines do things that would require intelligence if done by human [7]. Solutions built on AI technologies can help the users to detect fraudulent cases from the examples of previous fraudulent cases learnt by the system. It takes a machine learning approach that requires the understanding and manipulation of text as it deals with email communications. Text processing research uses Natural Processing Language (NLP) techniques to pre-process the text and extract the frequent scammed words known as the *keyword* [8].

Since to date, there is no existing research found to build a solution for pet scams, the research explores and studies research related to scammed email detection, the methods and results. Alireza et al. [12] categorized emails as scams, spams and hams as shown in Figure 2. Scams are defined as illegal emails, spams are the legitimate emails and hams are those emails exchange between users.

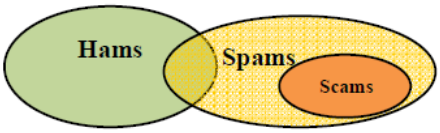

Figure 2: Email Categorization [12]

Among the popular methods in spam detection is to use the features of emails and frequency of spammed words as described by Pfleeger and Bloom [21]. This research is also referred as phishing email detection. Chandrasekaran et al. [22] classifies emails based on two structural attributes embedded in the email header: 1) Structure of email subject line and 2) Structure of email greeting in the email's body. Using a statistical method known as Support Vector Machine (SVM), the research attained an averaged on 85% accuracy on 5 series of experiments. Other methods include Logistic Regression (LR), Classification and Regression Trees (CART), Bayesian Additive Regression Trees (BART) and Artificial Neural Networks (ANN) as critically reviewed by Abu-Nimeh et al. [23]. There are existing tool for scammed email detection for both research and commercial use such as the PhishCatch [16] but none them explored the pet scams related emails.

This research adopted a non-binary classification of scammed emails by detecting the fraudulent patterns that are learnt in the collection of real scammed emails [9] and a simple mathematical calculation. After a successful training, our proposed system is able to label the emails into four categories – *Surely Scam, Scam, Potential Scam* and *Not Scam*.

3. PROPOSED METHODS

Data Collection

There is no publicly available dataset related to pet scams. Hence, we compiled our data based on the pet scammed emails published in a personal blogs, awareness websites and those that have been shared in forums by the victims. A total of 50 pet scammed emails have been compiled for our experiments. Additionally, to support the case study methodology, interviews are conducted with known real victims to learn their experiences and study the whole processes that resulted to monetary loss [20]. An example of the email with the potential scammed keyword based on frequency in other emails is shown in Figure 3.

Hello,
My name is kerry.
I am actually a mother of Luis(my lovely late 14 years old son) ,he passed away last two weeks in a car accident with my lovely husband.
I have weep until i do not know what else to do but to move along with life.

Actually i have 2 Maltese puppies (male / female) my husband bought for my late son some months ago.
They are now grown but my humble son whom i love so much just passed away and i am also disable due to the accident of which i am on a wheel chair ,so i really need a good home for the puppies.
I really do not have the strength and means to take care of the puppies.
Note: that I'm not selling this puppies, but offering for adoption. I want the best for the puppies, so in search of a pet lovely and caring home.
Please can you tell me more about yourself?
If interested, hoping to read from you.
Thanks.

Figure 3: An Example of Pet Scammed Email (Frequent Keywords are Underlined)

The interview discovered the following pet scammed chronology. The communications are done through email between scammers and victims.

1. Scammers shared brief story about the pets to shows how lovely the pets are.
2. Scammers post some lovely photos of the pets.
3. Scammers shared the story about the hardship of his or her family to grab the attention and sympathy of the victims.
4. Scammers asked the geographical details of the victim's location.

5. Scammers proposed for door-to-door shipping and will charge the victims for the 'shipping fee'.
6. Victims will bank-in the amount of money requested.
7. Victims never received any pet delivery.
8. Scammers' contacts are no longer accessible.

A full and original email communication is shown in **APPENDIX** for further reference.

System Architecture

The system architecture is shown in Figure 4. It consists of several components. Our solution is proposed as an extension to the web browser which will appear as an icon in the user email's toolbar. It serves as an additional feature to the ordinary email to automatically filter the pet scammed emails. The Pet Scammed Detector filtering engine is built on Visual Basic and link to the list of keywords database. This database compiles all potential per scammed keywords extracted from the collection of pet scammed emails.

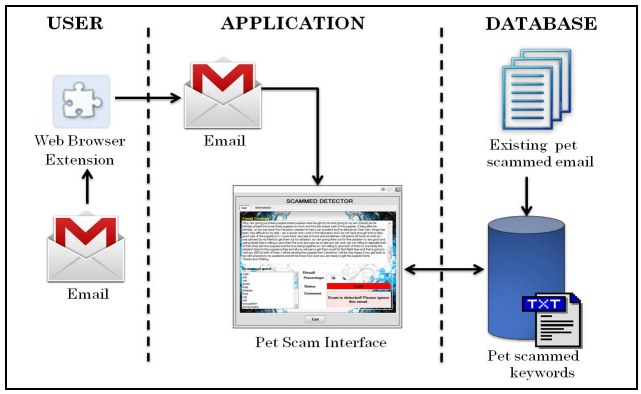

Figure 4: System Architecture

The flowchart of our proposed system is depicted in Figure 5. The raw text in email is pre-processed using standard NLP techniques, tokenized into vector of words before they get compared with the keywords in our database. Matched keywords are then calculated for its frequency values and compared against some threshold. The thresholds to identify and classify scammed email are based on the techniques that have been explained briefly in the next section.

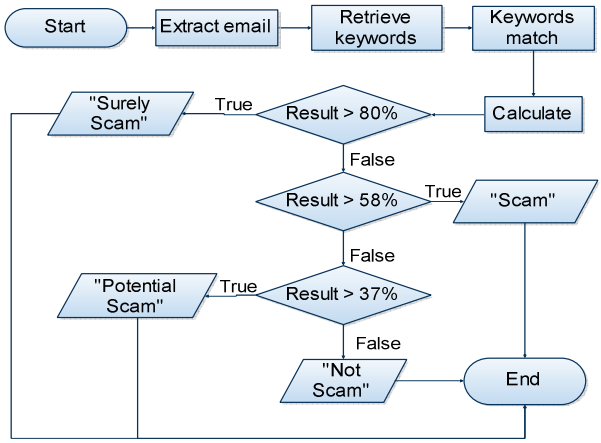

Figure 5: System Flowchart

Classification Methods

The collection of 50 emails is automatically sifted and frequency of each word is counted for each email. The result generated a maximum and a minimum frequency of the pet scammed keywords which are 32 and 5 respectively. The number between maximum and minimum frequency of those keywords is considered the range of the data set using Equation 1.

$$Range = Max - Min \quad (1)$$

We classified the emails into four different categories - *Surely Scam, Scam, Potential* Scam and *Not Scam.* The *Range* obtained in Equation 1 is divided by four, which is the number of email categories as shown in Equation 2.

$$Size = Range / Categories \quad (2)$$

The statistical information of pet scammed email collection is summarized in Table 1.

Table 1: Characteristics of Pet Scammed Email Collection

Total Email	50
Maximum Keyword	32
Minimum Keyword	5
Range	27
No. of Categories	4
Size	6.5

The four categories are ranked based on the frequency score values or the number of keywords extracted in an email as introduced in [9]. The frequency distribution table is shown in Table 2.

Table 2: Frequency Distribution of Pet Scammed Email Collection

No. of Keywords	Categories	Lower Boundary	Upper Boundary	Midpoint	Percentage of Categorization (%)
6 – 11	Not Scam	5.5	12.5	8	< 38
12 – 18	Potential Scam	12.5	19.5	15	> 37 and < 59
19 – 25	Scam	19.5	26.5	22	> 68 and < 81
26 - 32	Surely Scam	26.5	32.5	29	> 80

The percentage of the email categorization is calculated using Equation 3.

$$\textit{Percentage of Categorization} = [(Keyword + 1) * 100] / Max. \quad (3)$$

The example of the calculation for the highest percentage of *Not Scam* category is exemplified below:
Keyword = 11
Max = 32
Percentage of Categorization = [(11 + 1) * 100]/ 32 = 37.5%

We have compiled 66 common keywords related to pet scams from the 50 collected emails. Some of the keywords in our database are shown in Figure 5. The database considers all the variation of words for every lemma (root word). For example, the word die has four variations – *die, dies, died* and *dead*. This is important as the classification is highly dependable on the string similarity of the keyword search. Lemma has been experimented as the best form for document retrieval as shown by Zukerman and Raskutti [15].

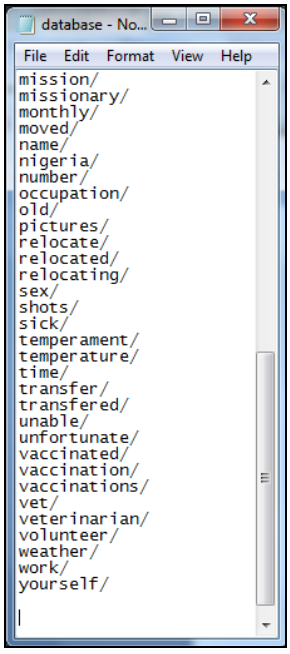

Figure 5: Pet Scam Keywords

4. EXPERIMENTAL RESULTS

The system prototype is built using Visual Basic 2010 Express Edition tool and the interface is shown on Figure 6. It is proposed to appear as an extension or additional feature to the current email interface. Once a suspicious scammed email is identified, user will click on the Pet Scam Detector button to get the email classified. An alert will be displayed to user with different colours to show the warning level of the potentially scammed email/

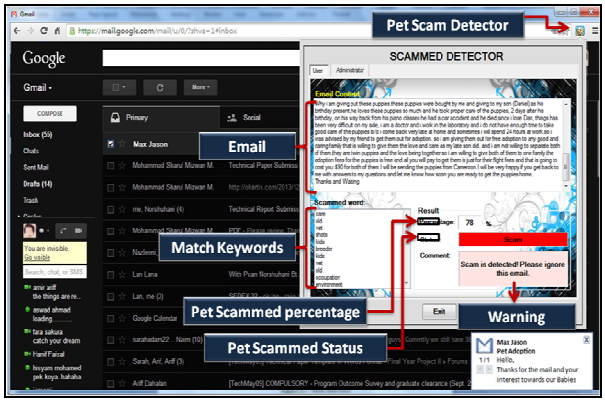

Figure 6: Pet Scammed Detector Interface

The system is evaluated using 50 unseen emails which include both the pet scammed and legitimate emails. The results are shown in Table 3.

Table 3: System's Generated Email Categorization

No. of Keywords	Category	No. of Emails
1-11	Not Scam	5
12-18	Potential Scam	21
19-25	Scam	17
26-32	Surely Scam	7

These automated generated results are then compared with the built-in string similarity matching algorithm in Microsoft Words and manual string similarity matching by five human experts for its accuracy. Partially, the results for five emails and their comparison are shown in Table 4. The shaded cells represent inaccurate results.

Table 4: Accuracy Test Results

DETECTOR	E1			E2			E3			E4			E5		
	W	%	S	W	%	NS	W	%	NS	W	%	SS	W	%	NS
Pet Scam Detector	12	67	/	4	22	/	4	22	/	18	100	/	6	33	/
Microsoft Word	12	67	/	4	22	/	4	22	/	18	100	/	6	33	/
Human 1	11	61	/	4	22	/	4	22	/	17	94	/	6	33	/
Human 2	12	67	/	4	22	/	4	22	/	18	100	/	5	28	/
Human 3	12	67	/	4	22	/	4	22	/	18	100	/	6	33	/
Human 4	12	67	/	4	22	/	4	22	/	18	100	/	6	33	/
Human 5	13	72	/	5	28	/	4	22	/	18	100	/	6	33	/
RESULT	71		100	86		100	100		100	86		100	86		100

Legend:
E1-E5 = Email 1 to Email 5 respectively.
W = Matched Keyword
S = Scam
NS = Not Scam
SS = Surely Scam

The overall results are visualized in Figure 7. The accuracy is 86%. The 14% inaccuracy rate was contributed by human errors in classifying the emails manually. The nature of pet scammed emails to be very lengthy and linguistic thus making the manual matching procedure with the long list of keywords in the database a cumbersome task.

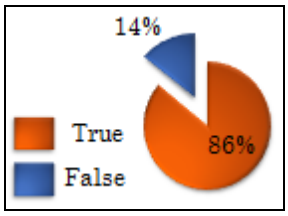

Figure 1: Total Accuracy for 50 Emails

5. CONCLUSIONS

The paper presents a technique to classify scammed email related to pet scams. The classification method has been successfully demonstrated on a Visual Basic application as an extension in the email application. The accuracy test has shown that the system is able to categorize scammed emails at fairly accurate rate. Our future development includes the detection of legitimate web contents that deal with pets. This will require the automated checking of the domain name address of the website, the validity of the posted images on the websites and the existence of the contact numbers.

6. REFERENCES

[1] “Scam Detector”. [Online]. Available: http://scam-detector.com/. [Accessed 11 December 2013].

[2] C. Simmons, "Losses in Internet scams hit new high," 4 March 2008. [Online]. Available: http://usatoday30.usatoday.com/news/washington/2008-04-03-4070959121_x.htm. [Accessed 11 December 2013].

[3] “Is the UK a Nation of Pet Lovers?," 2013. [Online]. Available:http://www.bret.org.uk/nationofpetlovers.html . [Accessed 9 December 2013].

[4] "U.S. Pet Ownership and Shelter Population Estimates," 27 September 2013. [Online]. Available: http://www.humanesociety.org/issues/pet_overpopulation/facts/pet_ownership_statistics.html#.UqcEU9IW3x0. [Accessed 9 December 2013].

[5] "Pet Statistics," 2013. [Online]. Available: http://www.petnet.com.au/pet-statistics. [Accessed 9 December 2013].

[6] "NZ Survey Reveals A Nation Of Pet Owners And Animal Lovers," 15 August 2011. [Online]. Available: http://www.wspa.org.nz/latestnews/2011/NZ_survey_a_nation_of_animal_lovers.aspx. [Accessed 9 December 2013].

[7] M. Negnevitsky, Artificial Intelligence - A Guide to Intelligence Systems, New Jersey: Addison Wesley Educational Publisher Inc, 2005.

[8] G. G. Chowdhury, (2003). Natural language processing. *Annual review of information science and technology*, *37*(1), 51-89.

[9] M. S. A. H. Maryam Mokhtari, A Novel Method in Scam Detection and Prevention using Data Mining Approaches. *Proceedings of IDMC2008*, 1-11.

[10] A. Alexander, "Internet Scam Statistics 2012 [Infographic]," 12 June 2012. [Online]. Available: http://ansonalex.com/infographics/internet-scam-statistics-2012-infographic/. [Accessed 10 July 2013].

[11] C. Simmons, "Losses in Internet scams hit new high," 4 March 2008. [Online]. Available: http://usatoday30.usatoday.com/news/washington/2008-04-03-4070959121_x.htm. [Accessed 11 December 2013].

[12] A. Saberi, M. Vahidi, & B.M. Bidgoli, Learn to Detect Phishing Scams using Learning and Ensemble Methods. In Web Intelligence and Intelligent Agent Technology Workshops, 2007 EEE/WIC/ACM International Conferences on (pp. 311-314). IEEE.

[13] "Scam Report," [Online]. Available: http://www.ipata.org/useruploads/files/scams/Scams_Report_April-30-2013.pdf. [Accessed 11 July 2013].

[14] "Looking for a Pet Online? Don't Get Scammed," [Online]. Available: http://www.ipata.org/pet-scams/. [Accessed 11 July 2013].

[15] I. Zukerman, & B. Raskutti, B. (2002, August). Lexical query paraphrasing for document retrieval. In *Proceedings of the 19th international conference on Computational linguistics-Volume 1* (pp. 1-7). Association for Computational Linguistics.

[16] W. D.Yu, S. Nargundkar, & N. Tiruthani, (2009, July). Phishcatch-a phishing detection tool. In *Computer Software and Applications Conference, 2009. COMPSAC'09. 33rd Annual IEEE International* (Vol. 2, pp. 451-456). IEEE.

[17] "U.S. Pet Ownership and Shelter Population Estimates," 27 September 2013. [Online]. Available: http://www.humanesociety.org/issues/pet_overpopulation/facts/pet_ownership_statistics.html#.UqcEU9IW3x0. [Accessed 9 December 2013].

[18] "Pet Statistics," 2013. [Online]. Available: http://www.petnet.com.au/pet-statistics. [Accessed 9 December 2013].

[19]"NZ Survey Reveals A Nation of Pet Owners and Animal Lovers," 15 August 2011. [Online]. Available: http://www.wspa.org.nz/latestnews/2011/NZ_survey_a_nation_of_animal_lovers.aspx. [Accessed 9 December 2013].

[20] R. Johansson, (2003, September). Case Study Methodology. In *Proceedings of the International Conference Methodologies in Housing Research* (pp. 22-24).

[21] P. Pfleeger, S. Bloom, “Canning Spam: Proposed Solutions to Unwanted Email”, Security & Privacy Magazine, IEEE, vol.3, no.2, Mar-Apr 2005, (pp. 40-7).

[22] Chandrasekaran, M., Narayanan, K., & Upadhyaya, S. (2006, June). Phishing Email Detection based on Structural Properties. In *NYS Cyber Security Conference* (pp. 1-7).

[23] Abu-Nimeh, S., Nappa, D., Wang, X., & Nair, S. (2007, October). A Comparison of Machine Learning Techniques for Phishing Detection. In *Proceedings of the Anti-phishing Working Groups 2nd Annual eCrime Researchers Summit* (pp. 60-69). ACM.

APPENDIX

Email 1:

Thanks for your mail regarding my kittens , they are still available to any one prepared to provide them with enough care and love , where they will be well spoil with all their need. they are 12 weeks old and are very good with kids and other pets. Very playful love to play around with toys and kids , they are health guaranteed and registered , vet checked . They are up to date on all their shots and will be coming along side with health papers and vet records.

The reason why i am giving them out to any lovely and caring home is because working hours are too much for me and it gives me no time to take care of them and moreover , they use to belong to my late daughter who just died in a car accident some months back . Being a single mum who just lost her only child , it has really been hard for me to take care of them and offer them the love they used to have . Due to all this and advice from some friends i work with , i have decided to give them out for adoption to a family where they will be well taken care of and treated with much love and care as my daughter use to do .

Since you are interested in getting them , i am willing to give them to you but you have to first assure me that you will take good care of them and that you will send their pictures to me monthly so that i can see how they are doing in their new home.
i will like to ask you some question because i believe it is correct to know the family wanting to adopt these lovely kittens . I do hope you can answer me the following:
-do you have kids?
-are you a breeder?
-do you have other pets?
-where are you located?
-how soon do you want them?
-How can i be sure you will take proper care of them?
-Will you be interested in the male or female or both ?

I sincerely hope you are not angry with all the questions i asked . I am just doing this because i really want these kittens to get into good hands . So many people are seeking to adopt them but i need to be confident on whom am given them to .
Please note that the above questions are important and thus you must answer every one of them . If not i am afraid i cannot let you have these kittens . So please tell me more about yourself and family.

Thanks and have a great day.

Email 2:

Thanks for keeping in touch, and for accepting one of my kittens as part of your family. I will look for a cheap pet delivery agency, that will transport and safely deliver the kittens to your door step. For safety and any unforeseen eventuality, I will register the kittens in a reliable agency that will issue a valid paper of proof during The movement .So do not worry the kittens will be shipped safely to your home.

These kittens are very friendly and get along with other household pets They are very good indoor and out door animals. they also love cuddles and will be great with your kids. As the distance keeps us apart, We will be entitled to use a pet delivery service you will pay only for the delivery to you which will cost you **800MYR** So that the kittens will be transported and safely deliver to your home address.

I will register the kittens for delivery to your home address and I will take care of the registration fees and also the transfer of ownership papers from my names to you name okay. So, you will pay only **800MYR** for the delivery of the kittens.

Note that you will have to pay this money directly to the agency and they will transport the kittens safely to your home address. I was told the agency will use the money for the following:
-To buy the flight ticket
-vaccination and medical checkup
-approval and transfer of ownership certificate
-quarantine and crate.
so all this must be validated before they can do the
delivery to you So you will have to pay the 800MYR to them before
delivery can be done to you

Also i was told by the agency that , after i have placed the kittens with them , the agency will contact you through phone and email to inform you that there kittens are with them , for you to confirm your address details (**since they will be doing a home delivery to you**) ,how to do the payment (**since you will have to pay 800MYR to them before delivery is done to you**) and finally they will provide you with a tracking code with which you will have to track the kittens online at their website to be sure that the kittens are with them and awaiting delivery to you. So if you are okay with all this kindly get back to me with the following information:Your full name?
Your home address?
Your phone number both house number and mobile number ?
Your ZIP code?
Your City?

With this information I will process the transfer of ownership papers of the kittens. I will register the kittens soonest for delivery to you .Since I have already prepared for the kittens for delivery I was told that if I register the kittens at the agency,
they will deliver the kittens the same day. Because it will be transported by an airplane with a few hours delivery .

I will let you know what the kittens needs later ,the kittens will be coming along with the following basket with what to eat, toys and clothing their food menu.. Also do promise to send me update of pictures with you and the kittens and also send me one when you receive the kittens.