ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 5, Issue 2, August 2015

Developing a Spam Email Detector

Maha Adham Bayati, Saadya Fahad Jabbar
Department of Computer Science, College of Science, Al-Mustansiriyah University - Iraq

Abstract*— Email is obviously important for many types of group communication that have become most widely used by millions of people, individuals and organizations. At the same time it has become a prone to threats. The most popular such threats is what is called a spam, also known as unsolicited bulk email or junk email. To detect spams, this work proposes a spam detection approach using Naive Bayesian (NB) classifier, where this classifier identifies email messages as being spam or legitimate, based on the content (i.e. body) of these messages. Each email is represented as a bag of its body's words (features). To catch up with the spammers latest techniques, a robust, yet up-to-dated dataset CSDMC2010 spam corpus (last updated 2014) - a set of raw email messages, was considered. To best perform, NB's environment was integrated with a list of 149 features proposed to include those commonly used by most spam emails. CSDMC2010 dataset was used to train and test NB classifier. Certain preprocessing (Tokenization, stop word removal, and stemming) was needed to drop out any redundant data. To further reduce dimensionality of the feature space, information gain (IG) and word frequency (WF) methods of feature selection were successfully tested against the email messages. Certain criteria were used to evaluate the performance of the proposed approach. To upraise accuracy level, certain statistics were suggested to extend NB algorithm with.

***Index Terms*— spam, spam filter, Naïve Bayesian classifier, Information Gain, Word Frequency.**

I. INTRODUCTION

A spam is an unsolicited (unwanted) email that the user does not want to receive and has not asked to receive; hence it is not wanted to be in one's inbox [1]. Spam emails have become a serious problem on the internet that the whole industry and individuals are suffering from the effects of this trouble. Tracing spammers has becomes a difficult and time consuming task owing to technological obstacles with email infrastructure. A spammer makes it more difficult and complicated by forging or hiding the origin of their messages. Even if spammers were traced, it is still difficult to take legal actions against them due to the internet decentralized architecture that has no central authority [2]. For the foresaid reasons, spam filtering has become of great importance and necessity. Generally speaking, spam filter is an application that classifies the incoming email messages as being a legitimate or spam emails using some particular method.

This paper proposes the spam detection approach to identify email messages as being spam or legitimate based on the content of these messages. Section two presents some of the published researches which cover the problem of a spam. Section three presents a background on spam resolutions and Naïve Bayesian classifier. Section four presents the methodology of the proposed work. The evaluation of results is discussed in section five. The last section presents the conclusions.

II. RELATED WORK

There are several publications concerning spams, each of which establishes certain approach or methodology. These are: [3], used different techniques in designing the spam filters including: K- Nearest Neighbor, Naïve Bayes, Bayes Additive Regression Tree, and Support Vector Machine (SVMs). In this research a real life data set was created from emails received by ten different email accounts and setup on different email servers. It was noticed that accuracy could be improved when the number of features and dataset size are increased. Results gained with Naïve Bayesian were better than those of other filters. [4], this work used a Bayesian, ANNs, K-NN, AIS (Artificial Immune System), SVMs, and rough sets classification. This study held a comparison between classifiers performance on the Spam Assassin spam corpus. In addition to the body of an email message, some fields of the header like the (from) and (subject) were also considered. The experiment was carried out by investigating the words that are frequently used in spam emails; 100 of these words were selected as features, The results with Naïve Bayesian were the best. [5], used three commonly machine learning techniques including (NB), (SVM) and Artificial Neural Network (ANN). The data set used was UCI spam-base dataset. The proposed spam filters showed the following performance evaluation: (ANN: Accuracy=93.44), (NB: Accuracy= 58.27), (SVM: Accuracy= 92.50), and finally (Combined: Accuracy = 93.94).The ANN achieved a slightly higher accuracy than linear kernel based SVM. While NB classification gave minimum accuracy with UCI spam-base dataset. [6], a spam classifier using machine learning algorithms including NB, SVM, and KNN was also proposed. The dataset used was spam assassin and the numbers of features used was 100 feature. The performance evaluation recorded for the three classifiers was: (NB: Acc = 99.46), (SVM: Acc = 96.90), (KNN: Acc = 96.20). On the contrary of the previous studies, NB gave a satisfying performance among the other learning methods.

III. BACKGROUND

1- Spam *Resolution*

There are many methods to resolve the spam problem. These methods are classified into non technological solutions, and technological solutions.

1-1 Non Technical Resolutions

As the name entails, there are methods that do not use any kind of technological tools to address the problem; they are

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 5, Issue 2, August 2015

only based on the reaction of the recipients. A non-technological solution demands the users and companies to take actions, which would frighten people from sending spams. They provide good attractive ideas with lots of difficulties to implement, but if an appropriate knowledge and fidelity is created on the side of email users, it will give good results. This type includes: Recipient Revolt, Customer Revolt, Vigilante Attack, Hiding the email address, and Contract-Law and limiting trial accounts Resolutions [7].

1-2 Technical Resolution

The technical resolutions are reactive in nature. This means, whenever a spam is presented at the user account, these techniques automatically starts to deal with the spam to eliminate it. This kind of solutions work towards making the spammer's job more difficult rather than stopping them from sending there unwanted messages [8]. There are many types of these techniques. Email filtering is the most popular one and seems to be the most effective way to fight spams at once. There are different types of filtering methods are:

1-2-1 Reputation–Based Filtering

This type of filtering methods based on information outside the email message content. These filters evaluate the reputation of one or more of the subscribers (recipient, sender, and Mediators). The subject of reputation calculation differs from one method to another. For example some filters use IP address while others use send or even sender domain [9]. This type of filtering method include: Blacklist Filter, Whitelist, Challenge/Response Systems, and Origin Diversity Analyzing Technique

1-2-2 Content–Based Filtering

As the name indicates, these filters based on examining the nature of the email content. Many end users use the content based filters to detect spams. Most of these filters try to understand the content of a message through some predictive keywords that are used to identify spam emails [10]. This type of resolutions includes:

a. Heuristic Filters: They are the ones that depend on looking for patterns which are commonly identified in spams. These patterns may be certain words, malformed message headers, phrases, capital letters, and exclamation marks [9]. One of the popular heuristic filters are the Rule-Based Filters. These techniques are based on critical words occurrence to classify email messages. An email message is then classified as "spam" if it has statistical features which are very close to those features of the unsolicited emails than to those of the legitimate emails [11].

b. Statistical Spam Filters: statistical filtering technique is one of the types of machine learning techniques for spam filtering. Machine learning technologies are employed in statistical filters not only to make anti-spam statistical filters responding quickly to the changes of spam without administrative intervention, but also to improve the performance of identifying spam [12]. The techniques applied in these filters are: Naïve Bayesian classifier, support vector machines, chi squared, maximum entropy models, Boosting, and memory-based learning techniques.

2- Naïve ***Bayesian Classifier (NB)***

This is a well-known practical probabilistic classifier that has been employed in many applications. One of these applications is the spam filtering. It is considered as one of the supervised learning algorithms for text classification. Bayes classifier depends on Bayes theorem, and the adjective naive is obtained by assuming that the features of a dataset are independent mutually. In other words, it is assumed that all attributes (features) of the training examples are independent of each other given the context of the class [13].

The NB classifier represents each pattern X (email) as an n-dimensional vector of feature values [f1, f2, …,fn] Given that there are l classes C_1, C_2 ,..., C_l, (where l = 2 spam and non-spam). The classifier expects an unknown pattern X belongs to a class that has the maximum posterior probability conditioned on X, i.e., X is assigned to class C_i if and only if

$P(C\ i|\ X) > P(C\ j|\ X)$ Eq (1)

For $1 <= j < i$ and $j \neq i$.

$P(C\_i|X)=(P(X|C\_i)P(C\_i))/(P(X))$ Eq (2)

However, to reduce the computational expenses involved, the classifier makes the naive or simplified assumption that the features (whose total number is denoted by n) are conditionally independent of one another. The independence class-conditional can be expressed as:

$$P\ (X|C\_i) = \prod_{j=1}^{n} P(f_j|C_i) \quad \text{Eq (3)}$$

As P(X) is a constant for each class, and P (C_i) = (|C_i |)/N, NB classifier needs to maximize only the P (X|C_i). This highly reduces the computation costs, since it counts the class distribution only. Bayesian classifier is very simple, requiring only a single scan of the data, thereby providing high accuracy and speed for large DBs [14].

IV. METHODOLOGY

The proposed system is a step forward towards building a spam filter with essential considerations to support performance in an on–line working environment. It aims to detect spam emails. For that, the Naïve Bayesian (supervised-learning) approach has been adopted to develop the proposed spam classifier.

a- Reading (.eml) files: This is a first step to be taken in the filtering system execution whenever a file with .eml extension is encountered. In this process, email parsing is done by breaking up its content into header and body. The body of the email is the only part that is considered in the proposed mechanism in order to be passed to the following process.

b- Preprocessing: Its main objective is to drop out any redundant data, hence keeping those only parts of an email body that give useful information which helps guiding efficient classification. Preprocessing consists of the following stages:

- Tokenization: Typically, the term "tokenization" referred to the process of breaking out a stream of text into its

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 5, Issue 2, August 2015

constituent meaning full units (called tokens). From a linguistic perspective, tokenization is accomplished on a word level, hence breaking the stream of text out into words. A predefined set of separators is used for this purpose. In the proposed tokenizer the punctuations mark, numbers, white space, and html tags were considered in breaking out the body of the email message into a sequence of features.

- Stop Words Removal: Stop words such as "the", "are", "with", "and", "to" etc. are some of the common words which occur almost in all textual information. These need to be removed because these stop words does not carry any useful information for helping to decide whether a mail message belongs to a classify or not. The elimination of the stop words from the mail message, will lead to a reduction in the dimensionality of the features space. The list of 429 stops words used in this thesis. This stop word list is obtained from [15].
- Stemming: is the term used in linguistic morphology and information retrieval to describe the process of reducing derived words to their word stem (base or root form). Many algorithms have been developed for stemming. Amongst are: look up algorithm, matching algorithm, and the popular Porter's stemming algorithm. For the proposed work, a look-up-like method was developed. In this method a table including the stem of candidate words together with corresponding derivatives is built. Part of this table is shown in Table 1. Note that data resource for this method was the Webster dictionary of English usage.

Table 1 Example of Words with their Stems

Root	DW 1	DW 2	DW 3	DW 4	DW 5	DW 6	DW 7
Actor	Actors	Actress	Actresses				
Address	Addresses	Addresser	Addressor	Addressable			
Adult	Adultly	Adultness	Adultery				
Bank	Banking	Banks	Bankable	Banker	Bankroll	Bankrupt	
Business	Businesses	Businessman	Businesswomen				
Free	Freer	Freest	Freeing	Freed	Frees	Freely	Freeness

c- Features Extraction: Is a method of construct the combinations of the variables to get around a problem which still describe the data with appropriate accuracy. A feature can be a word, an html tag, a phrase, etc. Since the "body" is the part of an email message this work is concerned with, and this body is a stream of textual data, and since the words of that body are considered to underline feature, of an email, a Bag of Word method was used throughout this work to carry on feature extraction. In this method, an email is represented as a bag of its body's words (features). A list of 149 features was considered, including those commonly used by most spam email. Some of these features were thought fully adopted from [16]. Additional others were suggested such as: "http", "www", and all strings of capital letters.

d- Features Selection: Two algorithms of feature selection were used in this work: Information Gain (IG) and Word Frequency (WF), both of which aimed at reducing dimensionality of the features. IG for individual feature F is calculated as in the equation (4).

$$IG(F) = -\sum_{j=1}^{k} p(C_j) \log p(C_j) + p(F) \sum_{j=1}^{k} p(C_j / F) \log p(C_j / F) + p(\bar{F}) \sum_{j=1}^{k} p(C_j / \bar{F}) \log p(C_j / \bar{F}) \quad \text{Eq(4)}$$

C j stands for the jth class out of total of K classes that are restricted to 2 (in the problem in hand): spam, non-spam, P (C j) stands for the probability of class C j, P (F) stands for the probability of a feature F being appeared in email messages, p(C j / F) stands for the probability of a feature F being appeared at least once in message of class C j, ($\bar{F}$): stands for the probability that feature F not being appeared in the training dataset, p (C j / $\bar{F}$): stand for the probability of feature F not being appeared in any of the messages of class C j [17].

WF: The simple word Frequency for a (feature, class) pair was defined by [18] as the number of messages in class C containing feature F, as shown in Equation(5).

$$WF = \sum_{j=1}^{k} pr(Cj) \; n \; C \; j \; F \quad \text{Eq (5)}$$

C j stands for the jth class out of total of K classes that are restricted to 2 (in the Problem in hand): spam, non-spam, pr (C j) stands for the probability of class C j, n C j F is the number of messages in class C j that has at least one occurrence of F.

e- Training: This is the process to train the proposed spam detection for doing the required classification on input emails. In this work 3800 emails from CSDMC2010 spam corpus were used for Training. This process was accomplished by stepping through the following Sequence of tasks:

1-present the NB classifier with each of the following sets of features:

Set-A: The original set of 149 features.

Set-B: Set of those only features selected according to IG Method.

Set-C: Set of those only features selected according to WF Method.

2- As a result, three classification models would consequently be obtained whose training results were to be used in testing phase:

All-feature NB model (NB-All F): the classifier model trained by set-A.

IG-based NB model (NB-IG F): the classifier model trained by set-B.

WF-based NB model (NB-WF F): the classifier model trained by set-C.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 5, Issue 2, August 2015

3- Calculate the priori probability and the conditional probability:

Compute the priori probability of each class p (Cs), p (Cns).

Where: P (Cs): is the probability of spam mails in the training dataset.

P (Cns) is the probability of non-spam mails in the training dataset.

Compute the conditional probability P (xi|Cj) for the email's value xi of its ith feature that exists in either of the foresaid sets A-C. This probability estimates the relative frequency of mails in Training dataset having value xi as the ith feature in class Cj.

F- Testing: This is the process to classify the unlabeled messages (in the dataset) into Spam and non-spam Mails. This process uses the result from training phase with entries of new unlabeled messages to classify it. At this step, a set of testing emails consists of 500 labeled emails was used. Conditional probabilities concerning email and posteriori probability were calculated at the testing process.

Now, it's time to continue calculating the conditional probability, but this time for the testing emails at each class, using the equation (3) and assuming the conditional independence of features. To avoid obtaining a result of zero out of this equation, Laplacian estimator was introduced each time the frequency xi of the ith feature in the testing email in hand was equal to zero. Thus, increment frequency of xi by one, as well as increment frequency of both classes (back in the training dataset) by a value equals number of available values of the ith feature.

At last, the posteriori probability is left to be calculated for each unlabeled email X accord to equation (2). Two probabilities are produced: P (Cs |X); probability of X being a spam email, P (Cns |X); probability of X being a non spam email. Finally, classification is attempted by deciding the class with the maximum posteriori probability to label email X with accordingly.

g- Extension of Naive Bayesian Classifier (ENB): Now that the NB classifier done, testing phase showed that certain percentage of misclassification was still exist. Is due to the slight difference between spam and non spam probabilities recorded for some emails. To deal with this issue, an ad-hoc tolerance algorithm was proposed to tolerate left over of misclassification by NB algorithm. It depends mainly on some statistical measured to enhance the overall system behavior. Algorithm 1 presents the proposed steps. It is worth to outline here that the key (strength) impact of this extension lies on this algorithm handle the effect of feature individually on misclassified email. Wherse NB, by itself, were conceding the overall effect of features on tested email, hence weaken the effect of probability some of the vital features that otherwise support robust classification.

Algorithm 1 ***Naive Bayesian extension***

Input : ME-list, // list of misclassified emails
S // set of features which is either ALL_F, IG_F, or WF_F

Output : emails that has been classified with NB classifier

Begin :

For each feature in S **do**

Initialize SF, NSF to zero // SF is spam frequency, NSF is non spam frequency

For each email X in ME-list **do**

If class label of X is "spam" then

SF = SF + frequency (feature)

Else NSF = NSF + frequency (feature)

if SF > NSF then class label of X = 0 // 0 for email spam

Else class label of X = 1 // 1 for non spam

End For

End For

End

V. SPAM DETECTION EVALUATION

This section presents the results of evaluation of performance of this work for solving the spam problem. The criteria that were used to measure the performance of this work are: accuracy (AC), recall(R), precision (P), false positive (FP), false negative (FN), and the error rate (ERR). These criteria were calculated by using the equations 6, 7, 8, 9, 10, and 11 respectively.

$AC = ((S\rightarrow S) + (L\rightarrow L)) / ((S\rightarrow S) + (L\rightarrow L) + (S\rightarrow L) + (L\rightarrow S))$ (6).

$R = (S\rightarrow S) / ((S\rightarrow S) + (S\rightarrow L))$ (7).

$P = (S\rightarrow S) / ((S\rightarrow S) + (L\rightarrow S))$ (8).

$FP = (L\rightarrow S) / ((L\rightarrow S) + (L\rightarrow L))$ (9).

$FN= (S\rightarrow L) / ((S\rightarrow L) + (S\rightarrow S))$ (10).

$ERR= AC - 1$ (11).

Where:

S: spam, L: legitimate, $S\rightarrow S$: The total number of spam emails that are classified as spam emails, $L\rightarrow L$: The total number of legitimate emails that are classified as legitimate emails, $S\rightarrow L$: The total number of spam emails that are classified as legitimate emails, $L\rightarrow S$: The total number of legitimate emails that are classified as spam emails.

Four experiments were conducted. (Exp1) conducted on the NB with all features (149 features), (Exp2) conducted on NB using IG with 75% of features, (Exp3) conducted on NB using WF with 75% of features, and (Exp4) conducted on extension of naïve Bayesian (ENB) with 75% of features. Table 2 depicts the results of these experiments. The results of ENB are better than the results of NB-All F, NB-IG F, and NB-WF F in all evaluation measures (i.e., AC, R, P ,FP, FN, ERR).

Table 2 Results of NB with four Experiments

Measure	Exp1	Exp2	Exp3	Exp4
AC	90	90.4	91	100
R	83	85.4	86	100

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 5, Issue 2, August 2015

P	87.3	87.7	88	100
FP	0.061	0.063	0.061	0
FN	0.170	0.146	0.142	0
ERR	10	9.6	9	0

VI. CONCLUSION

This paper has proposed approach to detect unsolicited emails. To detect those messages, certain classifier is needed to identify incoming messages as being spam or legitimate. Deciding a suitable classification method to gives a rewarding performance (in on off-line working environment) is what the proposed work aimed at. The Naïve Bayesian classifier was used in this approach. To classify emails into spam or legitimate, NB classifier needs to be trained, and then tested. The size of 3800 emails for training give the best results. The size of feature set has an evident effect on the performance of the spam classifier. The result gained with proportion 75% of features space was better than those gained with 50% and 25% proportions, in terms of all evaluation measures. Feature selection techniques considerably affect the dimensions of the data and consequently the computation time to construct the classifier. Despite the error rate that dropped down NB classifier at certain settings of its parameters, an ad-hoc statistical step (procedure) has been adequately able to fix those misclassifications. The accuracy of the proposed spam classifier has been affected by the dependency of features on the frequency of one to another's, when accuracy increased when all features are statistically independent of each other.

REFERENCES

[1] Thomas Choi, "Transactional Behaviour Based Spam Detection", master thesis, Department of Systems and Computer Engineering, Carleton University Ottawa, Ontario, Canada, April 2007.

[2] Jon Kågström, "Improving Naive Bayesian Spam Filtering", M.Sc. Thesis, Mid /Sweden University Department for Information Technology and Media, spring 2005.

[3] Tariq R. Jan, "Effectiveness and Limitations of Statistical Spam Filters", University of Kashmir, Srinagar, India, International Conference on New Trends in Statistics and Optimization, 2009.

[4] W.A. Awad, and S.M. ELseuofi, "Machine Learning Methods for Spam Email Classification", International Journal of Computer Science & Information Technology (IJCSIT), Vol 3, No 1, Feb 2011.

[5] Jincheng Zhang, and Yan Liu, "Spam Email Detection: A Comparative Study", Techniques for Data Mining Journal, December 6, 2013.

[6] S. Divya, and T. Kumaresan, "Email Spam Classification Using Machine Learning Algorithm", International Journal of innovative research in Computer and Communication Engineering, Vol.2, Special Issue 1, March 2014.

[7] Nouman Azam, "Comparative Study of Features Space Reduction Techniques

[8] for Spam Detection", M.Sc. thesis, National University of Sciences and Technology, 2007.

[9] David Ndumiyana, Munyaradzi Magomelo, and Lucy Sakala, "Spam Detection using a Neural Network Classifier", Online Journal of Physical and Environmental Science Research, Volume 2, Issue 2, pp. 28-37, April, 2013.

[10] Hasan. S. Alkahtani, Paul .G. Stephen, and Robert Goodwin, "A Taxonomy of Email Spam Filters", The 12th International Arab Conference on Information Technology, Naif Arab University for Security Sciences,11-12-2011.

[11] Hao Wu, "Detecting Spam Relays by SMTP Traffic Characteristics Using an

[12] Autonomous Detection System", Ph.D. Thesis, Loughborough University 2011.

[13] Valerio Freschi, Andrea Seraghiti, and Alessandro Bogliolo, "Filtering Obfuscated Email Spam by means of Phonetic String Matching", springer, Volume 3936, pp. 505-509, 2006.

[14] Hao Wu, "Detecting Spam Relays by SMTP Traffic Characteristics Using an Autonomous Detection System", Ph.D. Thesis, Loughborough University 2011.

[15] Sang-Bum Kim, Kyoung-Soo Han, Hae-Chang Rim, and Sung Hyon Myaeng, "Some Effective Techniques for Naive Bayes Text Classification", IEEE Transactions on Knowledge and Data Engineering, vol.18, no.11, pp.1457-1466, Nov. 2006.

[16] Sushmita Mitra, and Tinku Acharya, "Data Mining - Multimedia, Soft Computing, And Bioinformatics", Publisher: Wiley-Interscience; 1 edition September 25, 2003, ISBN: 978-0-471-46054-1.

[17] http://www.lextek.com/manuals/onix/stopwords1.html

[18] Serkan Günal, Semih Ergin, M. Bilginer Gülmezoğlu, and Ö. Nezih Gerek "On Feature

[19] Extraction for Spam E-Mail Detection", Springer, Volume 4105, pp. 635-642, 2006.

[20] Guo Qiang, "Research and Improvement for Feature Selection on Naive Bayes Text Classifier", IEEE Conference Publications, 2nd International Conference on Future Computer and Communication, vol.2, 2010, pp.156-159

[21] Wang Chunping, "The study on the spam filtering technology based on Bayesian algorithm" International Journal of Computer Science Issues, Vol. 10, Issue 1, No 3, January 2013.

AUTHOR BIOGRAPHY

1. Dr. Maha Al-Bayati

1. PERSONAL INFORMATION: E-mail: bayati_ma@yahoo.com

NATIONALITY: Iraqi

2. EDUCATION: Degree Institution Date B.Sc.. in Computer Science University of Technology 1987 M.Sc.. in Computer Science University of Technology 1992 Ph.D. in Computer Science University of Technology 1998 (*) Title of Ph.D. Thesis: " Genetic Algorithm-Based Robot Path Planner " (*) Title of Master Dissertation" Design and Implementation of a Parser for Verbal and Nominal Arabic Sentences"

3. ACADIMIC WORK EXPERIENCE: Title Institution From…To Assist. Prof. Department of Computer Science /Faculty of Information Technology/ Applied Science University / Amman /Jordan 2000-present Assist. Prof.

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 5, Issue 2, August 2015

Institute of Computer Technology /Tripoli / Libya 1998-2000 Assist. Prof. Department of Computer Science /University of Technology / Baghdad / Iraq 1998 Lecturer Department of Computer Science / University of Technology /Baghdad /Iraq 1992-1997 Part Time Lecturer Department of Computer Science / Al – Mansoure University (private) / Baghdad / Iraq 1996 Part Time Lecturer Department of Computer Science / Al – Moustansiria University / Baghdad / Iraq 1993

4. RELEVANT NON-ACADEMIC POSITIONS: Title Institution From…To Committee chair /Annual report on the faculty of information technology faculty of information technology/ Applied Science University / Amman /Jordan 2006-2007 Committee member / ETS for computer science faculty of information technology/ Applied Science University / Amman /Jordan 2006-2007 Committee chair /Students academic guidance / faculty of information technology faculty of information technology/ Applied Science University / Amman /Jordan 2006-2007 Training the staff of the computer center Institute of oil/ Ministry of oil/ Baghdad /Iraq 1995 System Designer Computer Center / University of Technology /Baghdad /Iraq 1987-1989

5.COURSES TOUGHT/ INTEREST: Data Structure and Algorithms in C++ , Data Structure and algorithm design in C, Operating Systems , Distributed Information Systems, System Software, Structured Programming , Computer Graphics, Artificial Intelligence, PROLOG ,Computer Skills, Visual Basic Programming , microprocessors and assembly language.

6. RESEARCH AREAS: Genetic Algorithms in Optimization and Machine Learning, E-Learning, Natural Language Processing, Speech Processing, Web Applications.

2. Saadya Faahad Jabbar

1. PERSONAL INFORMATION:

NATIONALITY: Iraqi

E-mail: Saadya _sf2009@yahoo.ca

2. EDUCATION:

Degree Institution Date

B.Sc. in Computer Science / University of Baghdad 2003

M.Sc. in teaching methods of computer / Higher Arab Institute for Educational and Psychological Studies / Iraq 2011

M. Sc. in computer science / University of Al Mustansiriyah 2015

3. RELEVANT NON-ACADEMIC POSITIONS

Manager of the electronic computer unit in collage of education inb rushed / University of Baghdad 2008-2012