International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169
Volume: 3 Issue: 1 351 - 359

Arabic Documents Classification Method a Step towards Efficient Documents Summarization

Hesham Ahmed Hassan
Faculty of Computer and Information Computer Science Department
Cairo University
h.hassan@fci-cu.edu.eg

Mohamed YehiaDahab
Faculty of Computer and Information, Computer Science Department
King Abdulaziz University
Mdahab@Kau.Edu.Sa

Khaled Bahnassy
Faculty of Computer and Information, Computer Science Department
Ain Shams University
khaled_elbahnasi@cis.asu.edu.eg

Amira M. Idrees
Faculty of Computer and Information, Information Systems Department
Fayoum University
ami04@fayoum.edu.eg

FatmaGamal
Faculty of Computer and Information, Computer Science Department
Cairo University
Cairo,11351, Egypt
Fatma_MCTC_2004@hotmail.com

Abstract

The massive growth of online information obliged the availability of a thorough research in the domain of automatic text summarization within the Natural Language Processing (NLP) community. To reach this goal, different approaches should be integrated and collaborated. One of these approaches is the classification od documents. Therefore, the aim of this paper is to propose a successful framework for agricultural documents classification as a step forward for a language independent automatic summarization approach. The main target of our serial research is to propose a complete novel framework which not only responses to the question, but also gives the user an opportunity to find additional information that is related to the question. We implemented the proposed method. As a case study, the implemented method is applied on Arabic text in the agriculture field. The implemented approach succeeded in classifying the documents submitted by the user. The approach results have been evaluated using Recall, Precision and F-score measures.

Keywords: *Classification, Natural language processing*

*****

I. INTRODUCTION

Document classification is a sophisticated problem confronting many areas of research such as information system and computer science [1]. Nowadays, data overload formulates a problem in categorizing useful documents from documents that are not of interest. This task is becoming a challenging task in many areas. The main task of classification is to assign a document to one or more classes or categories. In various actual scenarios, the capability to automatically classify a document into a fixed set of categories is extremely required, common scenarios involve classifying a huge volume of unclassified documents such as newspaper articles, scientific papers and legal reports. Many approaches have been proposed; we categorized them into three maincategories: Unsupervised Document

Classification, Supervised Document Classification and Semi-supervised Document Classification. In the following subsections we will give a review on these classification approaches.

The main objective of this paper is to create a framework for document classification that classifies a document submitted by the user into its adequate class; the classes that we used were imported from the Agrovoc thesaurus [2]. We combined the Naive Bayes classifier (NB) [3] together with regular expressions to fulfill this task. To calculate the priori probability of each class imported from Agrovoc we used the publications of the Central Lab for Agricultural Expert System (CLAES).

A. Unsupervised Document Classification

Unsupervised classification focuses on the idea of allocating categories to documents based only on their content without a training set nor predefined categories[4]. Unsupervised document classification is used to enrich information retrieval, being based on clustering hypothesis, which utters that, documents with related contents are significant to the same query [4]. A fixed group of text is clustered into groups that have similar contents. The similarity between documents is calculated with the associative coefficients; such as the cosine coefficient in the vector space model. Hierarchical clustering algorithms are mainly used in document clustering. The single link method is also used, as it is computationally reasonable, but the complete link approach seems to be the most effective though it is very computationally challenging [5]. Neural models are also used in implementing unsupervised document clustering [6].

B. Supervised Document Classification

In supervised document classification, approaches like Pattern recognition and machine learning are utilized to document classification. An example of these classifiers are neural networks [7], support vector machines [8], genetic programming [9], Various of these classifiers can be used in combination with unsupervised learning, i.e., unlabeled documents, but the accuracy of a classifier can be enhanced by using a small set of labeled documents [10]. The aim is to use a classifier which needs small amount of manually classified documents to be generalized.

C. Semi-supervised Document Classification

The use of semi-supervised document classification has emerged in the late 1990s [11]. The classification structure is someplace between supervised and unsupervised, where the category information is determined from the labeled data and the structure of the data from the unlabeled data [11].

II. BACKGROUNG

In the area of classification many remarkable work was presented. We will discuss in this section some of this work.Swales [12] defines a genre as “a class of communicative events, the members of which share some set of communicative purposes. These purposes are recognized by the expert members of the parent discourse community, and thereby constitute the rationale of the genre”. Swales has been criticized and made known to show the analyst with some amount of challenges [13]. Likewise, scholars from Critical Linguistics and Critical Discourse Analysis express of the ‘social activity’ linked to each genre [14]; they therefore prospect communicative purpose from a more socially-oriented perspective and make the identification of the social activity taking place central to genre identification. Investigators planning to afford guiding principle for genre analysis located their effort on how to control the analysis of a corpus of texts of the same genre, rather than on the criteria used to compile the corpus – genre identification has been broadly background knowledge of or about the ‘speech community’ who is using the genre [15].

Lee [16] presented complex feature selection in the framework of supervised genre classification. Their technique is based on identifying the terms that appear in many documents of a particular genre while being equivalently spread over topical classes, supposing that the genre-revealing terms should be independent of the topic. In their design, only the Bag-Of-Words model is used. The Bag-Of-Words model is effective for distinguishing between genres, particularly when used with stylistic features such as parts-of-speech and punctuation. Rather than achieving feature selection. There is a significant connection between the genre and a group of documents written in a similar design, and thus morphological features of the text has a significant function in distinguishing between the genres, as suggested in [17].

A lead analysis was led by Douglas Biber in the eighties [18]. He attempted to programmatically detect text types, which denote to groups of

documents corresponding to their linguistic subject like informational production or narrative concern etc., separately of their genre categories. Biber applied the multi-dimensional analysis technique using patterns of manually identified linguistic features, like tense or aspect markers or anaphora.

The naive Bayes classifier was effectively used in Rainbow text classification system [19]. The fundamental hypothesis of the naive Bayes is that for a given class, the probability of terms occurring in a document is independent of one another. When the volume of the training set is small, the term's frequency evaluations will not be sufficient; if a term doesn't appear in the training data set, its relative frequency will be zero. To solve this problem they applied the Laplace law of succession.

Li [20] used 'bag-of-words' document representation scheme (vector space model). They disregarded the structure of the document and the sequence of words in the document. The word-list in the training set comprises all the terms that emerge in the training models after excluding the stop-words (those words which are not necessary for retrieval, like 'the', 'some' or 'of') and the low-frequency words (which occur rarely in the training examples). A main obstacle of that model is the huge sparse matrix that results from it, which raises a problem of high dimensionality.

Rauber et al. in [21] presented genre clustering of documents according to a specified topic, using domain free features like frequencies of special characters, punctuation and stop-words. They utilized "self-organizing maps", a neural network learning model, for clustering the feature vectors. The target of Rauber's work is to integrate genres with the topic-based society of digital library. Genre clustering is accomplished only on topically coherent groups of documents. No inclusive analysis of the type of document clustering by genre is conducted.

Argamon et al. in [22] reviewed the allocations of unigrams, bigrams and trigrams of parts- of-speech, as well as pronouns and determiners, in the BNC corpus and revealed significant differences between non-fiction and fiction documents. Santini in [23] uses uni-/bi-/trigrams of parts-of-speech together with or without punctuation to create a supervised genre classification task on the BNC corpus. The part-of-speech n-gram model is not the best model for distinguishing genres in the BNC corpus.

Isa et al., in [24] used the Bayes formula to represent the document as a set of vectors according to a probability division revealing the categories that the document possibly will belong to. This probability distribution as the vectors to represent the document, the SVM is used to classify the documents. Guru et al., [25] extended Isa's work to represent documents using interval valued symbolic features. The distribution of terms probability in a document are used to develop a representation and is then used for classification purposes.

III. PROPOSED METHOD

The proposed method architecture is presented in figure 1, the main objective of the proposed method is to classify the document submitted by the user; the classes that we used were imported from the Agrovoc thesaurus [2]. We used the Naive Bayes classifier (NB) [3] to reach the target result. The first component extracts the classes and keywords from the Agrovoc thesaurus, the second component then creates a regular expression for each class and its keywords, the third component calculates the priori probability of each term imported from Agrovoc according to its existence in the publications of CLEAS. The fourth component calculates the posterior probability of these terms according to their existence in the document to be summarized.

Although NB classifier uses the independence assumption, the model is widely used in many applications such as text classification and information filtering (spam filtering) [26]. One of the major causes that NB model functions appropriately for text domain is that, the evidences are "vocabularies" or "words" showing in texts and the amount of the vocabularies is usually in the scale of thousands. The large size of evidences (or vocabularies) makes NB model work well for text classification problem [26]. Actually, it usually outperforms more complex classifiers such as Support vector machine (SVM) [27] or Relevance vector machine (RVM) [28], even when the underlying assumption of (conditionally) independent predictors is far from true. This advantage is especially pronounced when the number of predictors is very large.

In dealing with NB classifier, three issues should be kept in mind; first NB classifier requires a very large number of records to obtain good results [28]. Second, where a predictor category is not present in the training data, naive Bayes assumes that a new record with that category of the predictor has zero probability. These features will not affect the output efficiency of the proposed approach. With naive Bayes, however, the absence of this predictor actively "out votes" any other information in the record to assign a 0 to the target value (while, in this case, it has a relatively good chance of being a 1). In the proposed approach, the presence of a large training set helps mitigate this effect [28].

A. Extract Vocabulary Component

In the proposed approach, the classes and keywords were imported from the Agrovoc thesaurus [2] In the Extract vocabulary component; we extracted the classes, sub classes, and each sub class's keywords from the Agrovoc thesaurus and inserted them in system database. In the classification process we can predict the classification of an input text to the Agrovoc classes by observing the keywords. Generally, it is better to have more than one keyword to support the classification process. Typically, the more keywords we can gather, the better the classification accuracy can be obtained. We faced the problem of not having enough keywords in the Agrovoc to support each class, to solve this problem we had to manuallyadd keywords to some classes of the Agrovoc as an experimental study, these terms were added according to the expert. We also had to extract the terms thesaurus and the terms synonyms from our experts.

We have 238 classes imported from the Agrovoc for building our NB model; we faced some situations considering the keywords:

1-keywords that exist in the Agrovoc and don't belong to any class. According to expert's opinion, some of these keywords are added manually to its classes such as "إعياء" ("fatigue") which was added to class "النباتية الأمراض" (Plant diseases and disease management), other terms like "عظام" ("Pones"), were ignored.

إعياء	fatigue
عظام	Pones
تسحيب العظم	Pulling bone

Table 1 some of the terms don't belong to any class

2-Keywords that didn't exist in Agrovoc such as "مرض" which means ("disease"). And are considered effective in the classification process according to the expert opinion. These keywords are added manually to its class. For example the term "مرض" which means ("disease") is added to the class " الأمراض النباتية" (Plant diseases and disease management).

B. Create a Regular expression for each Vocabulary Component

In this component we programmatically created a regular expression for each of the vocabulary imported from Agrovoc. The regular expression will be useful in the matching process. In this component we created a regular expression taking into consideration each vocabulary's synonyms imported from our expert. This component was also responsible for providing some enhancements on the created regular expression some of the examples for these enhancements are as follows:

- Adding "ال" at the beginning of each term.
- Converting the "ي" or the "ى" at the end of each term to "ى|ي".
- The "أ" or "ا" or "آ" or "إ" at the beginning of each term to "ا|أ|آ|إ ".
- The "ه-" or "ة" at the end of each term to "ة-ه".
- Add optional spaces at the beginning and the end of each term.
- Add the names of the classes and subclasses to the term.

An example for the regular expression of the term " جيزة 177 " is " (((ال)?جيز[ةه] ? 177|جيز[ةه] 177 ?)) ". And the regular expression for the term " الأرز العطرى" is "((ال)?[ا|أ|إ]رز (ال)?عطر[ي|ى])" We then used the publications of (CLAES) center to give each term a weight by calculating term frequencies.

C. NB classifier Model component

The NB classifier model is divided into two phases, phase 1 is the priori probability phase and phase 2 is the posterior probability. The main aim of the priori probability is to give each term exported from the Agrovoc a priority in its class based on its occurrence in the publications of CLAES. The posterior probability is calculated after the user submits the document to be summarized; its aim is to find the Agrovoc class to which the document belongs. After determining the class to which the document belongs, we can determine which Agrovoc keywords we will be using in the sentence ranking component.

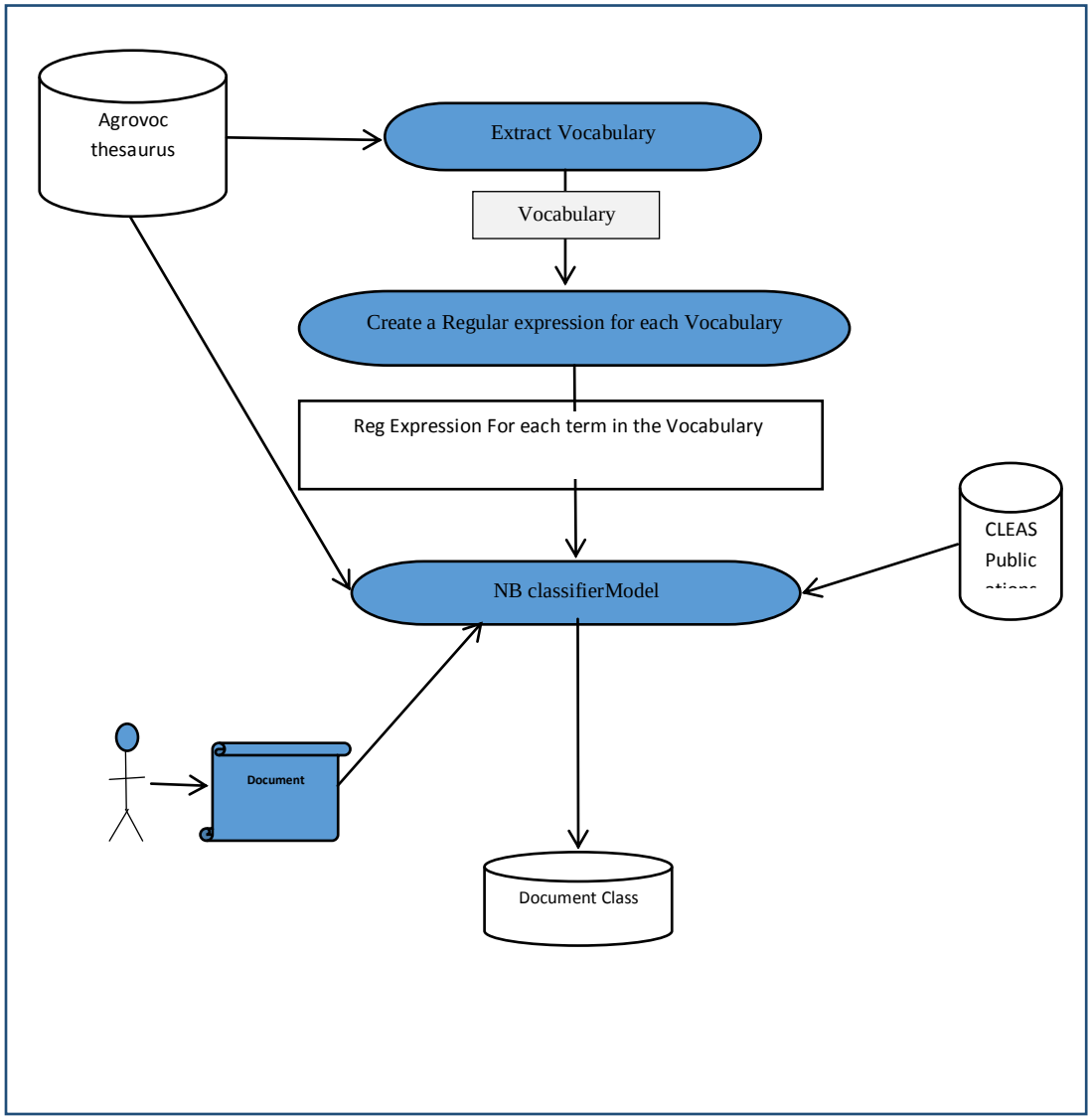

Figure 1: Classifier Component

Phase 1: The priori probability

Considering the data set that was imported from the publications of CLEAS, we had 2425 documents as training data set. And considering the terms and classes imported from the Agrovoc we have nearly 25000 terms and 238 classes (categories). Table 1 shows an example of the dataset that we have. The numbers are the frequency of the terms imported from the Agrovoc in the documents of CLAES publications. For example, as shown in table 2, the word محاصيل (Crops) occurred 14256times in class الأمراض النباتية(Plant diseases) and 253 times in التسميد (Fertilization) class.

Name	Fertilizer أسمدة	أعفان rot	Pests آفات	Splatter بقع	Proliferation تكاثر	Pollution تلوث	damages خسائر	Wilting ذبول	Irrigation ري	Fungi فطريات	Crop محاصيل	Disease مرض	Pathogens المرض مسببات	Pest Control مكافحة الآفات
الأمراض النباتية Plant diseases		2086	9139	27621			1320	9250		3960	14256	26061	71	753
التسميد Fertilization	1923					443			71995		253			
فسيولوجية النباتات – التكاثر Physiological plants - Reproduction					244									

Table 2: Sample of the classes of the Agrivoc and the frequency of some of the terms in the CLEASE publications.

For each Agrovoc term we calculate the term's priority using equation 2.

term priority = argmax $P(C_j) \prod P(X_i \mid C_j)$	Equation (2)

Where:
C_j are the classes imported from the Agrovoc thesaurus
X_i are the terms imported from the Agrovoc thesaurus
$P(C_j)$ frequency of class C_j
$P(X_i \mid C_j)$ is the frequency of term X_i within Class C_J
$P(C_i)$ is the priori probability of each class = number of documents in a class / number of all documents

The number of vocabularies in our database is 25755. The number of all documents is 2425. For example the class الأمراض النباتية in the database was

P (الأمراض النباتية) = 1583/2425 = 0.6527835051546392 ≈ 0.65

P (التسميد) = 392 /2425 = 0.1616494845360825 ≈ 0.16

N_I is the total number of word frequency of each class. In the database the N_I of class الأمراض النباتي was 70907 while the N_I of class التسميد was 77850.

$P(W_i \mid C_i)$ = the conditional probability of keyword w occurrence given a class c.

In the database the P (مرض | الأمراض النباتية) of class الأمراض النباتي ≈ 0.83.

Finally we can determine the document class by selecting the class with the highest weight and its super classes. The document class directs us towards the cue phrases and their weight which will lead us in the next system phase.

Phase 2: posterior probability

The posterior probability is calculated after the user submits the sample document. Figure 2 contains a sample paragraph.

Input Paragraph:

محصول الارز

يتعرض محصول الأرز للإصابة ببعض الأمراض ومن أهمها مرض اللفحة والتبقع البنى:

1- مرض اللفحة:

هو أشد أمراض الأرز خطورة و يتخذ شكلا وبائيا فى بعض السنوات فهو يصيب النبات فى جميع أطوار حياته .ففى طور النمو الخضرى يصيب الأوراق بحيث تظهر بقع صغيرة رمادية إلى زيتونية اللون محاطة بحافة بنية و تستطيل و تصبح مغزلية و فى حالة زراعة الصناف القابلة للإصابة تتشابك البقع مما يؤدى إلى جفاف الأوراق. فى طور السنبلة يصيب هذا المرض السنابل حيث يتلون عنق السنبلة بلون بنى و تصبح السنبلة فارغة تماما أو جزئيا وكذلك قد تكون الإصابة جزئية على فرع أو أكثر من فروع السنبلة ويؤدى ذلك إلى ضمور الحبوب على هذا الجزء وفى حالات الإصابة المتأخرة على السنابل يشاهد ضمور فى الحبوب مما يؤدى إلى نقص فى المحصول يتناسب مع ميعاد حدوث الإصابة على السنابل.

Rice crop

Rice crop is exposed to certain diseases the most important of which are the blight and brown spotting :

1blight :

Is the most severe disease of rice. It takes an epidemic form. In some years it infects the plant in all phases of its life. It develops bible vegetative growth that affects the leaves so that they appear small patches of gray to olive color surrounded by the edge of the structure swell and become spindled, in the case of cultivation of varieties susceptible intertwined spots, which leads to dehydrating the leaves .

In the process of spike, it affects the spike, the neck color turns brown, and the Spica is completely or partially empty, and may also be partially injured on one or more of the branches which leads to atrophy of the grain on this part and in the incidence of late atrophy in grain which leads to a decrease in yield commensurate with the time limit on the incidence of spikes.

Figure 2 : Input Text

To calculate the posterior probability we will take the following steps:

- The first step: To count the number of words which turned out to be 135
- Second step: Using regular expression, compare the vocabulary extracted from the document to that from the Agrovoc thesaurus.
- Third step calculate the term frequency for each term found in both the document and the Agrovoc thesaurus.
- Finally using this equation (argmaxP(Cj) ∏ P(Xi | Cj)), the class and sub class with the greatest value is selected.

Output: Main Class الأمراض النباتية(Plant disease)

Where:

Class is the name of the class according to the Agrovoc., Term is the terms imported from the Agrovoc and found in the document to be summarized. N_i is the total number of word frequency of each class. Class_Prior_priority is frequency of the class in the (CLAES) center publications. Term Frequency is the frequency of the term in the (CLAES) center publications. No_of_Documents is the number of the (CLAES) center publications in which the term was found. Number of document are the number of documents in which each term existed according to the class. P_W_C is the conditional probability of keyword occurrence given a class. Class_Temp_priority is the final priority of the class using the formula in the following equation

$$P(\boldsymbol{c_i}|\,W) = P(\boldsymbol{c_i}) \times \prod_{j=1}^{v} P(w_j \mid \boldsymbol{c_i})$$

Where $P(\boldsymbol{c_i}|\,W)$ is the posterior probability for each class.P(ci) = the priori probability of each class. P(wi | ci) = the conditional probability of keyword occurrence given a class

Note: To avoid the "zero frequency" problem, we applied Laplace estimation by assuming a uniform distribution over all words which is the total number of documents in the (CLAES) center publications.

lass	Term	Ni	Class Prior priority	Term Frequency	Number of Documents	P_W_C
الأمراض النباتية	أعفان	70907	0.65	2086	105	0.03
الأمراض النباتية	آفات	70907	0.65	9139	193	0.12
الأمراض النباتية	بقع	70907	0.65	27621	193	0.38
الأمراض النباتية	خسائر	70907	0.65	1320	153	0.02
الأمراض النباتية	ذبول	70907	0.65	9250	153	0.13
الأمراض النباتية	طرق التربية	70907	0.65	950	192	0.01
الأمراض النباتية	فطريات	70907	0.65	3960	154	0.05
الأمراض النباتية	محاصيل	70907	0.65	14256	181	0.19
الأمراض النباتية	مرض	28793	0.18	26061	121	0.83
الأمراض النباتية	مسببات المرض	28793	0.18	71	42	0.00
الأمراض النباتية	مكافحة الآفات	28793	0.18	753	121	0.02
التسميد	أسمدة	77850	0.16	1923	62	0.02
التسميد	تلوث	77850	0.16	443	49	0.01
التسميد	جودة	77850	0.16	738	49	0.01
التسميد	ري	77850	0.16	71995	62	0.90
التسميد	عناصر	77850	0.16	1834	60	0.02
التسميد	محاصيل	77850	0.16	253	49	0.00
فسيولوجية النباتات - التكاثر	محاصيل	244	0.01	0	0	0.00

Table 3: Naive Bayes classifier values for the text in figure 2

IV. Conclusion

This paper shows a method for agricultural Arabic documents classification as a step forward to finding a precise answers to user questions, it can be improved by exploiting summarization techniques to extract more than just the answer from the document in which the answer resides. This is done using a graph search algorithm which searches for relevant sentences in the discourse structure, which is represented as a graph. The experiment results showed success in most cases and it triggered some problems. First, we recommend using automatic class extraction to overcome the problem of insufficiency of data imported from different corpora, which will in turn minimize data storage.

Second, we recommend using an advanced NLP technique to recognize inclusive relations and to solve the problem of the hidden pronoun.

REFERENCES

[1] Document Classification for Newspaper Articles Dennis Ramdass&ShreyesSeshasai 6.863 Final Project Spring 2009 May 18, 2009

[2] Agrovoc Web Services-Improved, real-time access to an agricultural thesaurus, 2006, Boris Lauser,.Margherita Sini, Gauri. Salokhe,

[3] Caruana, R.; Niculescu-Mizil, A. (2006). "An empirical comparison of supervised learning algorithms". *Proceedings of the 23rd international conference on Machine learning*. CiteSeerX: 10.1.1.122.5901

[4] Van Rijsbergen, C.J. (1979). Information Retrieval. Buttherwords, London

[5] P. Willett. Recent trends in hierarchical document clustering: a critical review. Information Processing & Management, 24(5):577-597, 1988.

[6] MacLeod, K. An application specific neural model for document clustering. Proceedings of the Fourth Annual Parallel Processing Symposium, vol.1, p. 5-16, 1990

[7] Li, Wei; Lee, Bob; Krausz, Franl and Sahin, Kenan. Text Classification by a Neural Network. Proceedings of the 1991 Summer Computer Simulation Conference. Twenty-Third Annual Summer Computer Simulation Conference,p. 313-318, 1991.

[8] T. Joachims, Text Categorization with Support Vector Machines: Learning with Many Relevant Features. Proceedings of the European Conference on Machine Learning, Springer, 1998.

[9] Svingen, B. Using genetic programming for document classification. FLAIRS-98. Proceedings of the Eleventh International Florida Artificial Intelligence Research, p. 63-67, 1998.

[10] K. Nigam, A. McCallum, S. Thrun, & T. Mitchell. Text classification from labeled and unlabeled documents using EM, machine learning, 2000.

[11] Benkhalifa, M., Bensaid, A. and Mouradi, A. Text categorization using the semi-supervised fuzzy c-means algorithm. 18th International Conference of the North American Fuzzy Information Processing Society - NAFIPS, p. 561-5, 1999.

[12] J. M. Swales. Genre Analysis: English in Academic and Research Settings. Cambridge Applied Linguistics. Cambridge University Press, 1990.

[13] Askehave, I. (1999). Communicative Purpose as Genre Determinant. Journal of Linguistics, 23, 13- 2, 1999.

[14] Fairclough, N. Discourse and Social Change. Cambridge: Polity Press. (1992).

[15] Bhatia, V. K. (1993). Analysing Genre: Language Use in Professional Settings. London; New York: Longman.

[16] Y.-B. Lee and S. H. Myaeng. Text genre classification with genre-revealing and subject-revealing features. In Proceedings of the 25th International ACM SIGIR Conference, pages 145–150, 2002

[17] Y. Seki, K. Eguchi, and N. Kando. Analysis of multi-document viewpoint summariza- tion using multi-dimensional genres. In AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications, pages 150–153, 2004.

[18] D. Biber. Variation Across Speech and Writing. Cambridge University Press, 1988

[19] Gentle Introduction to RainBow. URL: http://www.cs.cmu.edu/afs/cs/project/theo-11/www/naive-bayes/gentle_intro.html

[20] Jain, A. K., and Li, Y. H. 1998. Classification of Text Documents. The Computer Journal, Vol 41, pp. 537 – 546 .

[21] A. Rauber and A. M¨uller-K¨ libraries. In Proceedings of the ACM/IEEE Joint Conference on Digital Libraries, pages 1–10, 2001. ogler. Integrating automatic genre analysis into digital

[22] S. Argamon, M. Koppel, J. Fine, and A. R. Shimoni. Gender, genre, and writing style in formal written texts. Text, 23(3), Aug. 2003 .

[23] M. Santini. A Shallow Approach to Syntactic Feature Extraction for Genre Classification. In Proceedings of the 7th Annual Colloquium for the UK Special Interest Group for Computational Linguistics, 2004.

[24] Isa, D., Lee, L. H., Kallimani, V. P., and Rajkumar, R. 2008. Text document reprocessing with the Bayes formula for classification using the support vector machine. IEEE Transactions on Knowledge and Data Engineering. Vol. 20, pp. 23 – 31.

[25] Guru D. S., Harish B. S., and Manjunath. S. 2010. Symbolic representation of text documents. In Proceedings of Third Annual ACM Bangalore Conference.

[26] Haruechaiyasak, C. &Damrongrat, C. (2008) "Article Recommendation Based on a Topic Model for Wikipedia Selection for Schools". The Eleventh International Conference on Asian Digital Libraries (ICADL 2008), pp.339-342.

[27] Cortes, C. &Vapnik, V. (1995). Support-vector network. Machine Learning,20, 1–25.

[27] M. E. Tipping. The Relevance Vector Machine. In S. A. Solla, T. K. Leen, and K.-R.M□uller, editors, Advances in Neural Information Processing Systems 12, pages 652{658. MIT Press, 2000.