DISSERTATION
On

FILTERING THE NEWS ITEM USING ID3 APPROACH

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE AWARD OF DEGREE
Of

MASTER OF ENGINEERING
(Computer Technology and Application)
Delhi University, Delhi

Submitted By:

YOGESH SHARMA

University Roll No 10075

Under the Guidance of:

Dr. Akshi Kumar
Assistant Professor

Department Of Computer Science and Engineering
Delhi College of Engineering, Delhi

DEPARTMENT OF COMPUTER ENGINEERING
DELHI COLLEGE OF ENGINEERING
(NOW DELHI TECHNOLOGICAL UNIVERSITY)
DELHI UNIVERSITY
2011

CERTIFICATE

I hereby certify that the work is being presented in the thesis report entitled, "**Filtering The News Item using ID3 Approach**", submitted by me in partial fulfilment of the requirements for the award of degree of Master of Engineering in Computer Technology & Application at Delhi College of Engineering, Delhi, is a authentic record of my own work carried out under the supervision of **Dr. Akshi Kumar** and refers other researcher's works which are duly listed in the reference section.

The matter presented in this thesis has not been submitted for the award of any other degree of any university.

(Yogesh Sharma)

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge.

Dr. Akshi Kumar

Assistant Professor

Department of Computer Science and Engineering

Delhi College of Engineering, Delhi - 110042

ACKNOWLEDGEMENT

No volume of words is enough to express my gratitude towards my guide **Dr. Akshi Kumar**, Assistant Professor, Computer Science and Engineering, Delhi College of Engineering, Delhi, who has been very concerned and has aided for all the materials essential for the preparation of this thesis report. She has helped me to explore this vast topic in an organized manner and provided me with all the ideas on how to work towards a research-oriented venture.

I am also thankful to **Dr. (Mrs) Daya Gupta**, Head of the Department, Computer Science and Engineering, Delhi College of Engineering, Delhi for the motivation and inspiration that triggered me for the thesis work.

I would also like to thank the staff members and my colleagues who were always there at the need of the hour and provided with all the help and facilities, which I required for the completion of my thesis work.

Most importantly, I would like to thank my parents and the almighty for showing me the right direction out of the blue, to help me stay calm in the oddest of the times and keep moving even at times when there was no hope

Yogesh Sharma
M.E. (Computer Technology and Application)
Department of Computer Science and Engineering
Delhi College of Engineering, Delhi-42

ABSTRACT

As the Web continues to grow, it has become increasingly difficult to read through all those information and go through to all the search pages to check which information is important and which is not till you find some meaningful information.

The information on the internet is so vast that if we want to collect and analyze any meaningful information or news about a particular subject, we need to have a system that will automatically analyze, sort and respond to only the relevant and important information while filter out the irrelevant information or news. There are various techniques and algorithms (such as machine learning) that come under the process of data mining, can be used for separating the relevant and irrelevant news from a dataset.

In this thesis, We have made the attempt to create a working news filter which will filters out the relevant and related news for the user out of the enormous amount of news results available for any searched word .

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATION

ABBREVIATIONS	EXPANSION
p(R|W)	Probability that the article is Relevant if the given word exists in it.
p(R'|W)	Probability that the article is Irrelevant if the given word exists in it.
P(R)	Probability that the article is relevant.
P(R'|W)	Probability that the article is relevant.
N	Total no. of words from the list that were found in the new result article
p(R|Wi)	Probability of the article being relevant given that a specific word was found in it.
i	Counter for calculating the product of the probabilities.
T	Relevant Article
F	Irrelevant Article
No.T	No of Relevant Articles
No.F	No. of Irrelevant articles

Table of Contents

Chapter 1

Introduction

In this chapter I will expand why we had chosen this work as my major thesis, how I motivated to do this work along with some issues regarding the work done earlier on the same and finally we will discuss how we accomplished the objectives in the contribution.

1.1 Motivation

The most popular way to look for news or any information on the Web is to use Web search engines such as Google. Many users begin their Web activities by submitting a query to a search engine. However, as the size of the Web is still growing and the number of indexable pages on the Web has exceeded eight billion, it has become more difficult for search engines to keep an up-to-date and comprehensive search index. Users often find it difficult to search for useful and high-quality information on the Web using general-purpose search engines, especially when searching for specific information on a given topic and if the user want the same news again he has to go back to search engine and again need to search for the same information or news that he searched long time back and again has to go through to all the pages to look for that information. So there has to be a system where we can store the searched news, filters out the relevant and irrelevant and can view that information whenever we can.

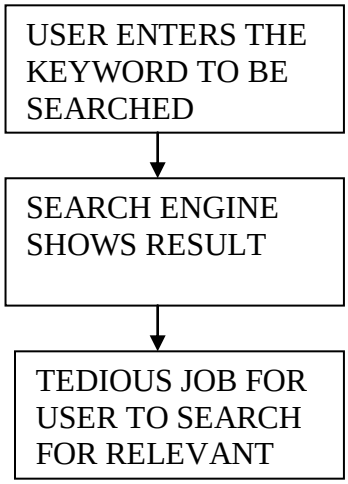


Figure 1.1 Present Scenario

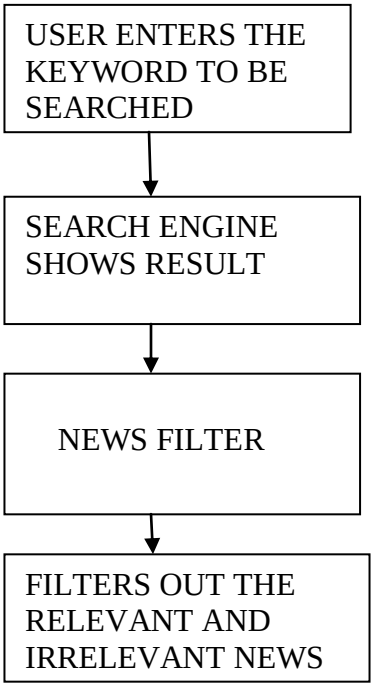


Figure 1.2 Modified Process

1.2 Research Objective

A search on a Search Engine returns a list of search results for a specified keyword or a phrase. Since the searched keyword or phrase may have several meanings and inferences, thus many unwanted results are expected to be present in the search results returned by the Search Engine. Even if the results displayed correspond to the correct meaning of the phrase, still many unwanted and irrelevant results are expected.

The search engine does not know which meaning or inference of the search keyword the user is interested in, so it returns the results for all the possibilities. This project works on developing a filter which sorts out only the relevant search results for the user out of the list of all the search results which are returned by the Search Engine.
For the implementation of the proposed filter we will take a dataset which is the input to the filter. A specific search keyword "CAS" (abbreviated form of 'Context Aware Services') has been considered for its initial implementation. Context-aware services are a computing technology which incorporates information about the current location of a mobile user to provide more relevant services to the user. An example of a context-aware service could be a real-time traffic update or even a live video feed of a planned route for a motor vehicle user. Context can refer to real-world characteristics, such as temperature, time or location. This information can be updated by the user (manually) or from communication with other devices and applications or sensors on the device. Apart from 'CAS' implying Context Aware Services, there are many other inferences or abbreviations of this word like: Chemical Abstracts Services, Central Authentication Services, Casualty Actuarial Society etc.

1.3 Proposed Work

After getting the initial results for the typed Search word in the search engine, the search results are taken from the search engine and we create a dataset of the articles given by the search engine, the user now trains the application through T (True) and F (False) in the dataset to convey which articles are relevant and which others are irrelevant for him. Using this input, the application in future displays only wanted and relevant filtered results to the user as depicted in the figure 1-2. The user can thus create training data sets for each search Keyword.

Since the application learns from experience, it is an intelligent application. The application can be trained even while it is displaying filtered results, which can further enhance the accuracy of the application

1.4 Organization of thesis

In the above chapter, we had discussed the motivation, problem statement and objective of the thesis. Chapter 2 provides the literature review of related works. Chapter 3 has the complete details design of the new approach, their algorithm, and limitations. Chapter 4 shows the experimental setup and results of the proposed system and finally chapter 5 consists of the conclusion and possible future work or directions in this area and it finally ends with the links and references (bibliography) and appendices.

1.5 Chapter Summary

In this chapter we had discussed the motivation of the author to do this work. The author had described the objectives of the thesis and proposed how to accomplish these objectives in his research using machine learning algorithms.

Chapter 2

Literature Review

In this chapter, we will provide a literature review for related work in this area. The use of machine learning algorithm for the purpose of news filtering (data mining) that came into existence a few years back. There are some machine learning algorithms that gained tremendous interest in recent years due to successful application of these algorithms on web mining or news filtering. We have made an attempt to organize some of related literature with respect to their relevance to this task. We hope the result of this literature review to help anyone who would like to design a new method for a different task based on what is known about previous methods.

2.1 Machine Learning

To solve a problem on a computer, we need an algorithm. An algorithm is a sequence of instructions that should be carried out to transform the input to output. For example, one can devise an algorithm for sorting. The input is a set of numbers and the output is their ordered list. For the same task, there may be various algorithms and we may be interested in finding the most efficient one, requiring the least number of instructions or memory or both.

For some tasks, however, we do not have an algorithm—for example, to tell spam emails from legitimate emails. We know what the input is: an email document that in the simplest case is a file of characters. We know what the output should be: a yes/no output indicating whether the message is spam or not. We do not know how to transform the input to the output.

Considered spam changes in time and from individual to individual. What we lack in knowledge, we make up for in data. We can easily compile thousands of example messages some of which we know to be spam and what we want is to “learn” what constitutes spam from them. In other words, we would like the computer (machine) to

extract automatically the algorithm for this task. There is no need to learn to sort numbers, we already have algorithms for that; but there are many applications for which we do not have an algorithm but do have example data.

There are few algorithms available in machine learning for data mining or web mining such as ID3 (Iterative Dichotomiser 3) and C4.5 which are also referred as decision tree algorithms of machine learning which we will going to discuss next in this chapter.

2.2 Decision Tree

Decision tree learning algorithm has been successfully used in expert systems in capturing knowledge. The main task performed in these systems is using inductive methods to the given values of attributes of an unknown object to determine appropriate classification according to decision tree rules. A decision tree is a tree in which each branch node represents a choice between a number of alternatives, and each leaf node represents a decision. Decision tree are commonly used for gaining information for the purpose of decision -making. Decision tree starts with a root node on which it is for users to take actions. From this node, users split each node recursively according to decision tree learning algorithm. The final result is a decision tree in which each branch represents a possible scenario of decision and its outcome.

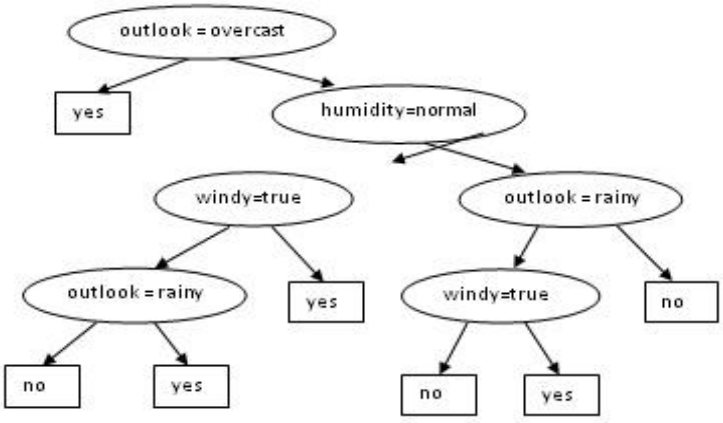


Figure 2.1 A Sample Decision Tree

2.2.1 ID3 - Basic Machine Learning Algorithm

ID3 is a simple decision tree learning algorithm developed by Ross Quinlan (1983). The basic idea of ID3 algorithm is to construct the decision tree by employing a top-down, greedy search through the given sets to test each attribute at every tree node.

Advantages of using ID3

1. Understandable prediction rules are created from the training data.
2. Builds the fastest tree.
3. Builds a short tree.
4. Only need to test enough attributes until all data is classified.
5. Finding leaf nodes enables test data to be pruned, reducing number of tests.
6. Whole dataset is searched to create tree.

Disadvantages of using ID3

1. Data may be over-fitted or over-classified, if a small sample is tested.
2. Only one attribute at a time is tested for making a decision.
3. Classifying continuous data may be computationally expensive, as many trees must be generated to see where to break the continuum.

2.2.2 C4.5 - Machine Learning Algorithm

C4.5 is an algorithm used to generate a decision tree developed by Ross Quinlan. C4.5 is an extension of Quinlan's earlier ID3 algorithm. The decision trees generated by C4.5 can be used for classification, and for this reason, C4.5 is often referred to as a statistical classifier.

C4.5 addresses the following issues not dealt with by ID3:

- Avoiding over fitting the data
    - Determining how deeply to grow a decision tree.

- Reduced error pruning.
- Rule post-pruning.
- Handling continuous attributes.
    - e.g., temperature
- Choosing an appropriate attribute selection measure.
- Handling training data with missing attribute values.
- Handling attributes with differing costs.
- Improving computational efficiency.

2.2.3 NB Tree - Decision Tree

The present invention provides a hybrid classifier, called the NB-Tree classifier, for classifying a set of records. According to the present invention, the NB-Tree classifier includes a Decision-Tree structure having zero or more decision-nodes and one or more leaf-nodes. At each decision-node, a test is performed based on one or more attributes. At each leaf-node, a classifier based on Bayes Rule classifies the records. Furthermore, the present invention provides a method for inducing the NB-Tree classifier from a set of labeled instances. To induce the NB-Tree classifier, a utility C_1 of a Bayes classifier at a root-node is first estimated. Next, a utility D_1 of a split into a plurality of child-nodes with a Bayes classifier at the child-nodes is estimated. The utility of a split is the weighted sum of the utility of the child-nodes, where the weight given to a child-node is proportional to the number of instances that go down that child-node. Next, it is determined if C_1 is higher than D_1. If C_1 is higher than D_1, the root-node is transformed into a leaf-node with a Bayes classifier. If C_1 is not higher than D_1, the root-node is transformed into a decision-node, and the instances are partitioned into a plurality of child-nodes. The method then recursively performs the previous steps for each child-node as if it is a root-node. The present invention approximates whether generalization accuracy for a Naive-Bayes classifier at each leaf-node is higher than a single Naive-Bayes classifier at the decision-node.

2.3 Chapter Summary

In this chapter we had elaborated the literature review or the research work, done on filtering the news items. All the research made is based on the decision tree algorithm of machine learning algorithm which includes ID3 algorithm and in the end we discussed about the decision tree which can also be made using Naive Bayes classifier called NB Tree.

Chapter 3

Design and Architecture

In this chapter we will discuss about the approach we are using in the designing of the system. The chapter also include the method that will be used for system designing. We also give the architecture of the system which gives us a brief idea about the procedure for the system

3.1 ID3 Decision Tree

ID3 is a simple decision tree learning algorithm developed by Ross Quinlan (1983). The basic idea ofID3 algorithm is to construct the decision tree by employing a top-down, greedy search through the given sets to test each attribute at every tree node. In order to select the attribute that is most useful for classifying a given sets, we introduce a metric---information gain.

To find an optimal way to classify a learning set, what we need to do is to minimize the questions asked (i.e. minimizing the depth of the tree). Thus, we need some function which can measure which questions provide the most balanced splitting. The information gain metric is such a function.

3.1.1 Entropy

In information theory, entropy is a measure of the uncertainty about a source of messages. The more uncertain a receiver is about a source of messages, the more information that receiver will need in order to know what message has been sent.

- Entropy(s) = - p+log2 (p+) -p-log2 (p-) for a sample of negative and positive elements.

- The formula for entropy is:

$$Entropy(S) = \sum_{i=0}^{c} Pi \log Pi$$ .....Equation 3.1

The Entropy of a sample being relevant or irrelevant is based on the probability of the word that exist in the sample or not thus it is important to find the probability of a word. These probability values are very significant in the determining the relevance of a new result Article.

The Counts are expressed as:

No.T = No. of Relevant Articles containing this word.

No.F = No. of Irrelevant Articles containing this word.

No.Total = No. of Total Articles containing this word.

This Probability is expressed as:

P (Relevant |Word) = p (R|W) =No.T/No.Total

This gives the Probability of an article being relevant if the given word exists in it.

There is another Probability to be considered:

P (Not Relevant |Word) = p (R'|W) = No.F/No.Total

This Probability is actually the negation of p (Relevant |Word) or p (R|W) .It is expressed as:

P (R' |W) = 1 - p (R|W)

3.1.2 Information Gain

Measuring the expected reduction in Entropy As we mentioned before, to minimize the decision tree depth, when we traverse the tree path, we need to select the optimal attribute for splitting the tree node, which we can easily imply that the attribute with the most entropy reduction is the best choice. We define information gain as the expected reduction of entropy related to specified attribute when splitting a decision tree node.

The information gain, Gain(S, A) of an attribute A:

Gain(S, A) = E (Current set) - ∑ E (all child sets).....Equation 3.2

We can use this notion of gain to rank attributes and to build decision trees where at each node is located the attribute with greatest gain among the attributes not yet considered in the path from the root. The intention of this ordering is to create small decision trees so that records can be identified after only a few decision tree splitting and to match a hoped for minimalism of the process of decision making.

With the advent of the Web and various specialized digital libraries, the automatic extraction of useful information from text has become an increasingly important research area in data mining. In this paper we discuss a new algorithm that extracts both the topics expressed in large text document collections and models how the authors of documents use those topics.

3.1.3 The ID3 Algorithm

ID3 Algorithm for Decision Trees

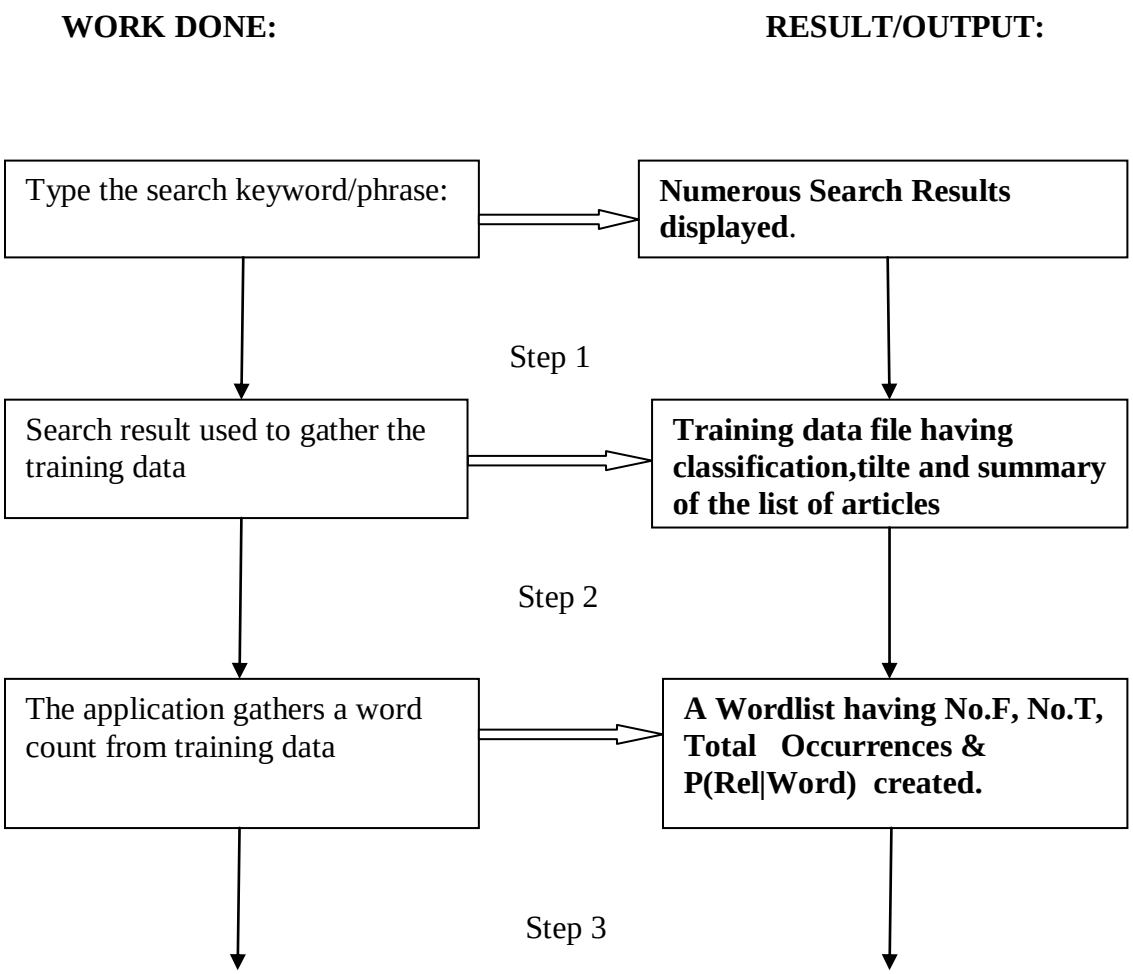
ID3 (Examples, Target_Attribute, Attributes)

- Create a root node for the tree
- IF all examples are positive, Return the single-node tree Root, with label = +
- If all examples are negative, Return the single-node tree Root, with label = -
- If number of predicting attributes is empty, then Return the single node tree Root, with label = most common value of the target attribute in the examples
- Otherwise Begin
    - A ← The Attribute that best classifies examples
    - Decision Tree attribute for Root ← A
    - For each positive value, vi, of A,
        - Add a new tree branch below Root, corresponding to the test A = vi
        - Let Examples(vi), be the subset of examples that have the value vi for A

- If Examples(vi) is empty
    - Then below this new branch add a leaf node with label = most common target value in the examples
    - Else below this new branch add the subtree ID3 (Examples(vi), Target_Attribute, Attributes – {A})

- End
- Return Root

3.2 Architecture of the System



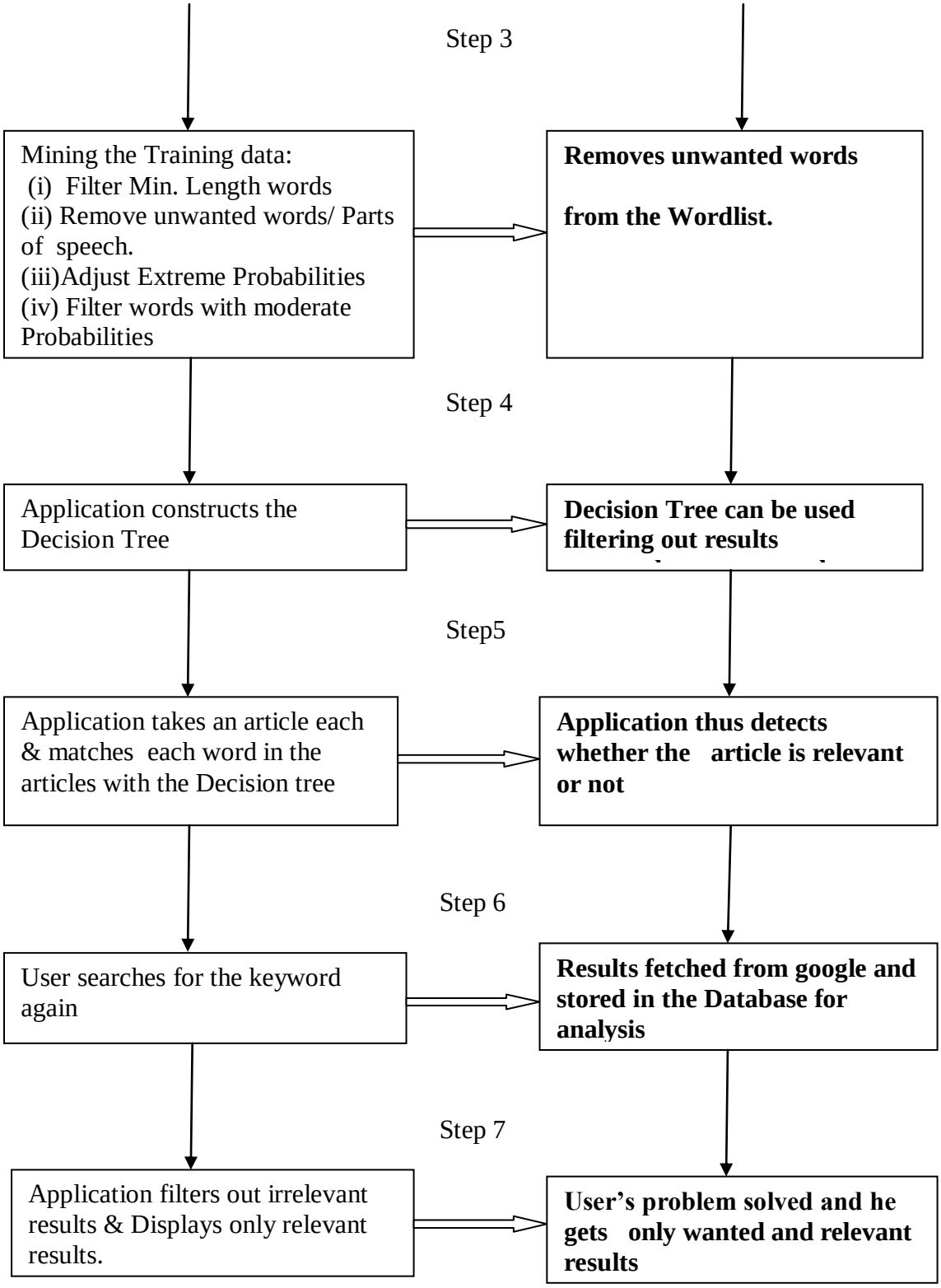


Figure 3.1 Architecture of the System

3.3 Chapter Summary

In this chapter, we had discussed in detail about the approach used to build a system, which is the ID3 machine learning algorithm. We also discussed about the functions, which plays important role in the ID3 algorithm, i.e. the Entropy and the Information gain. Finally, the chapter ends up on the architecture for the system which gives a complete overview of the system and its approach.

Chapter 4

Implementation

In this chapter we will implement the design proposed in the previous chapter using various datasets like sample data, training data, dataset for stop words, datasets for punctuation marks, datasets for the word counts and their probabilities and the dataset for the word count with attributes.

While doing implementation we will also see how the data will be trained and what are the steps required for the mining of the data to be trained.

4.1 Mining the Training Data

The training data is not very accurate and thus it cannot be directly used in the algorithm in determination of the relevance of the articles. It contains a lot of inconsistencies and inappropriate words and values.

The word list in the training data contains a lot of parts of speech, numbers, signs and symbols which do not actually reflect the relevance of an article. It contains words like: is , am ,are ,which, such, rather, suppose, although, other, they, ? , !, 2011, February etc. Existence of such words sometimes gives very inaccurate values of probabilities
The mining process was divided into 4 stages:

4.1.1 Filtering words of minimum length

We only consider words in the training data which have a minimum length. Here, words of length less than *2* characters were filtered off by a filter. The application takes this threshold value as input from the user. A function takes in this value and automatically filters off the words with minimum length.

4.1.2 Removing inappropriate words/ symbols

All the parts of speech, dates, numbers and symbols were removed.
Following parts of speech were removed from the training data:
Articles – E.g.: is, the, a etc.
Prepositions – E.g.: above, over, through etc.
Conjunctions – E.g.: and, rather, though, although etc.
Pronouns – E.g.: am, are, they, we, them etc.
Interjections- E.g.: eek, eh, encore, eureka, yeah!
Adjectives – E.g.: good, amazing, bad, small etc.
Adverbs – E.g.: almost, just, newly, nearly etc.
Special Characters: , . / " : ; { [}] | \ ~ ` ! @ # % ^ & * ( ) _ - + =
Dates: Months: January, February, November, December .

4.1.3 Adjusting the extreme Probabilities:

During the process of finding the words from the wordlist and their p(R\W), many words have this value equal to 0 or 1. These numbers, although impressive, are potentially over-fitting. Take for instance an article including *Russia*. According to the algorithm, this article is about mining. The training data also shows that all articles with Mining = T are classified as relevant. But a Russian singer may one day perform at the Hollywood Palladium. Therefore $p(T \mid Mining) < 1$.

Performance can be part of the sentence: "The Palladium price performed well last night." Therefore, the probabilities have been slightly adjusted as
If p(R|W) = 1 , then adjust p(R}W) to 0.99.
If p(R|W) = 0 , then adjust p(R}W) to 0.01.

4.1.4 Removing words having probabilities in a specific range.

Many words in the list have probabilities near 0.5 which do not give a clear indication whether the article is relevant or not if that word exists in the article or not.
The words which occurred often and most occurrences are of the same type (i.e. the empirical p(Relevant | Word) is either very high or low) were specifically considered. Rest of the words having moderate values of p(Relevant | Word) were filtered off.

Here the words which satisfy the below condition pass through the filter and rest are filtered off.

$$(0< p(\text{Relevant} \mid \text{Word})< 0.4) \quad \text{OR} \quad (0.6< p(\text{Relevant} \mid \text{Word})< 1.0)$$

4.2 Technology Used

The system which is proposed in chapter 3 is based on machine learning paradigm and we had also learnt the algorithms of machine learning which is the basic ID3 algorithm, a decision tree. For the implementation we have used the following tools:

FRONT END

Language Used for Interface: Microsoft Visual Studio Solution with Visual Basic Express 2008

BACK END

Here we have taken a news data that I have searched form a search engine and stored the searched result in the notepad which will act as database for my system

4.3 Datasets

As discussed in chapter 1, a sample news data of "CAS" i.e. 'Context Aware System' has been considered for its initial implementation. Supposedly, the user is interested in all the results corresponding to the abbreviation 'Context Aware Services". Apart from 'CAS' implying Context Aware Services, there are many other inferences or abbreviations of this word like: *Chemical Abstracts Services,* Central Authentication Services, and Casualty Actuarial Society etc.

Now the data may inconsistencies like stop words, punctuation marks, numbers and different symbols, which have to be removed from the dataset which gives a reliable and more accurate training data.

These four datasets are further elaborated as given below in the next section.

4.3.1 Dataset for Stop Words

S.No	Stop Words	S.No	Stop Words
1	a	21	plus
2	an	22	regarding
3	the	23	round
4	aboard	24	save
5	about	25	since
6	above	26	than
7	across	27	through
8	after	28	to
9	against	29	toward
10	along	30	towards
11	amid	31	under
12	among	32	underneath
13	anti	33	unlike
14	around	34	they
15	as	35	this
16	at	36	those
17	before	37	us
18	behind	38	we
19	below	39	what
20	per	40	whatever

41	which	86	nor
42	whichever	87	but
43	who	88	or
44	whoever	89	yet
45	whom	90	so
46	whomever	91	after
47	whose	92	although
48	you	93	as
49	yours	94	As if
50	beneath	95	As long as
51	beside	96	hey
52	besides	97	hi
53	between	98	hmmm
54	beyond	99	is
55	but	100	are
56	by	101	how
57	concerning	102	when
58	considering	103	where
59	despite	104	will
60	down	105	would
61	during	106	shall
62	except	107	should
63	excepting	108	was
64	excluding	109	were
65	following	110	can
66	for	111	could
67	from	112	la
68	in	113	my
69	inside	114	your
70	into	115	mine
71	like	116	yours
72	minus	117	uh
73	near	118	absentmindedly
74	of	119	adoringly
75	off	120	nothing
76	on	121	one
77	onto	122	One another
78	opposite	123	other
79	outside	124	others
80	over	125	ours
81	past	126	ourselves
82	yourself	127	several
83	yourselves	128	she
84	for	129	some
85	and	130	somebody

131	someone	176	everyone
132	something	177	everything
133	that	178	few
134	theirs	179	he
135	now	180	her
136	soon	181	hers
137	still	182	herself
138	then	183	him
139	today	184	himself
140	tomorrow	185	with
141	weekly	186	within
142	when	187	without
143	yesterday	188	all
144	abroad	189	another
145	anywhere	190	any
146	away	191	anybody
147	everywhere	192	anyone
148	here	193	awkwardly
149	In	194	beautifully
150	inside	195	briskly
151	out	196	brutally
152	outside	197	carefully
153	somewhere	198	cheerfully
154	there	199	competitively
155	underground	201	eagerly
156	upstairs	202	effortlessly
157	extremely	203	extravagantly
158	not	204	girlishly
159	quite	205	gracefully
160	rather	206	grimly
161	really	207	lazily
162	terribly	208	lifelessly
163	too	209	loyally
164	very	210	quietly
165	after	211	quickly
166	afterwards	212	quizzically
167	annually	213	really
168	before	214	recklessly
169	daily	215	remorsefully
170	never	216	ruthlessly
171	until	217	savagely
172	up	218	sloppily
173	upon	219	so
174	versus	220	stylishly
175	via	221	unabashedly

222	As much as	265	themselves
223	As soon as	266	these
224	As though	267	oh
225	because	268	ouch
226	before		
227	By the time		
228	Even if		
229	Even though		
230	if		
231	In order that		
232	In case		
233	lest		
234	once		
235	Only if		
236	Provided that		
237	since		
238	So that		
239	than		
240	that		
241	though		
242	till		
243	unless		
244	until		
245	when		
246	whenever		
247	where		
248	wherever		
249	while		
250	well		
251	ah		
252	alas		
253	dear		
254	eh		
255	er		
256	hello		
257	hullo		
258	his		
259	I		
260	it		
261	its		
262	itself		
263	anything		
264	both		

Table 4-1 Stop Word

4.3.2 Dataset for Punctuation Marks and Numbers

We will remove all the punctuation marks, special symbols and the numbers.

S.No.	Punctuation Marks
1	','
2	'.'
3	':'
4	' '
5	'\n'
6	'\0'
7	'\r'
8	'-’
9	','
10	‘+'
11	'='
12	'*'
13	'&'
14	'^'
15	'%'
16	'$'
17	'#'
18	'@'
19	'!'
20	'~'
21	'\'"
22	'\"
23	'\\'
24	'1'
25	'2'
26	'3'
27	'4'
28	'5'
29	'6'
30	'7'
31	'8'
32	'9'
33	'0'

Table 4-2 Punctuation Marks

4.3.3 Training Dataset for the News on 'CAS'

For the implementation we have taken the news on the 'Context Aware Services' from a search engine and then select which news is relevant and which one is not. The relevant news is prefixed by T (True) and irrelevant is prefixed as F (False).

F;Automating Capcom Using Mobile Agents And Robotic Assistants? We have developed and tested an advanced EVA communications and computing system to increase astronaut self-reliance and safety ...

F;Using mobile agents as roaming security guards to test and improve? This paper discusses the design and implementation details of MAST (Mobile Agent-based Security Tool), a new mobile agent-based network security approach ...

F;Optimized Wireless Web Browsing Using Mobile Agents? The use of wireless data comunications is increasing rapidly despite limited bandwidth and reliability ...

F;Security of Mobile Agent in Ad hoc Network ? In a very simple form a Mobile Agent is an independent piece of code that has mobility and autonomy behavior ...

F;Mobile agents: basic concepts, mobility models, and the Tracy toolkit? Students and researchers can use the book as an introduction to the concepts and possibilities of this field and as an overview of ongoing research ...

F;Facilitating Information Sharing Using Mobile Agents? Interests in Mobile Agents (MA) are rising because there are several benefits which can be associated with their employment ...

F;USING MOBILE AGENTS FOR OFF-LINE COMMUNICATION AMONG MOBILE HOSTS? In dynamic networks such as packet radio and ad-hoc wireless network, each node acts as a mobile router ...

F;Secure Internet Applications Based on Mobile Agents? The increasing importance of the Internet has motivated the exploration of new execution models based on mobile and dynamic entities to overcome the limits of the client/server model ...

F;CAS | Jasig Community? Welcome to the home of the Central Authentication Service project, more commonly referred to as CAS ...

F;Computer algebra system - Wikipedia, the free encyclopedia? A computer algebra system (CAS) is a software program that facilitates symbolic mathematics ...

T;Mobile Agents: Can They Assist with Context Awareness? This position paper argues that the mobile agents paradigm is a useful and important technology enabling pervasive and ubiquitous computing ...

T;Context-Based Addressing: The Concept and an Implementation for Large-Scale Mobile Agent Systems? We introduce the notion of context-based addressing, i.e. the ability to refer to and send messages to a collection of agents based on their current context, without knowing the precise identities of the agents ...

T;Context Aware Mobile Commerce Using Agent Technology - IEEE? Advances in e-commerce have resulted in significant progress towards strategies, requirements, and development of e-commerce applications ...

T;Ubiquitous Computing? Mobility is a major reason of dynamics in a system ...

F;Casualty Actuarial Society: Home? The Casualty Actuarial Society is a professional organization of actuaries whose purpose is the advancement of the body of knowledge of actuarial science ...

F;CAS | Trade Promotion Management | Optimization | Retail Execution? CAS, the industry leader in Trade Promotion Management Optimization solutions for Retail Execution of consumer packaged goods ...

F;Home | CAS? Welcome to the Centre for Aviation Studies (CAS) at University of ...

F;Computer Aided Services? CAD Conversion Services of high quality. Computer Aided Services, India delivers accurate Paper to CAD, Architectural Visualization and Offshore ...

F;Centre for Atmospheric Sciences, Indian Institute of Technology Delhi? The Centre for Atmospheric Sciences has a multidisciplinary team of highly qualified meteorologists, oceanographers, physicists, applied mathematicians ...

T;A MOBILE AGENT-BASED COMMUNICATIONS MIDDLEWARE FOR DATA STREAMING? In this paper we introduce the FlexFeed framework in the context of military combat operations. FlexFeed realizes the notion of Agile Computing for streaming data communications and implements a flexible, robust and efficient

publish/subscribe infrastructure for dynamic ad hoc environments under resource and policy constraints ...

F;CAS? Below you will find a list of discussions in the CAS forums at the India Broadband Forum ...

F;Logo for CAs? Logo for CAs. ICAI CA Shiksha eLearning. Members. Know Rules & Regulations ...

T;Context Provisioning and SIP Events? There is a general consensus that future services in mobile networks will be user-tailored and adaptive to the user's needs. Since context awareness can provide the required means for creating such services, it has become an important topic in research and also in industry ...

T;CONTEXT-AWARE TELECOMMUNICATION SERVICES? Telecommunication Services, Context, Context Awareness, Middleware, Routing, Addressing, Messaging, Screening ...

T;Commodity markets mostly climb, oil nears 80 dollars? MSN Malaysia News - On the London Platinum and Palladium Market, platinum climbed to 1541 dollars an ounce ... Palladium increased to 460 dollars an ounce from 456 dollars. ...

T;An Ambient Intelligence Application Integrating Agent and Service? This paper presents an agent-based approach into a more general service oriented architecture for addressing the requirements of accessibility content and services in an ambient intelligence context. The developed agent-based information system provides infomobility services for the special requirements of mobility impaired people...

T;AUGMENTING MOBILE SERVICES USING THE SEMANTIC WEB?

The Semantic Web has recently attracted the attention of both researchers and practitioners in the information systems field. In this paper we explore an application of Semantic Web in mobile context. In particular we focus on the development of advanced models of mobile service provision by contextualizing users' interaction characteristics through an upper level annotation ontology ...

T;Context addressing using context-aware flooding? Due to the proliferation of small networked mobile devices, the number of (indirectly) interconnected services in pervasive computing environments may grow without bound. The network contains a potentially

enormous amount of context aware services that sense, gather and distribute context information ...

F;CAS (Code access security)? This article first starts with the basic concepts of CAS like evidence, permission, code groups and caspol.exe ...

4.3.4 Test Dataset for search keyword "CAS"

F;CERN Accelerator School? The CERN Accelerator School holds training courses for accelerator physicists and engineers twice a year ...

T;Context addressing using context-aware flooding - OMNeT++ Network? Due to the proliferation of small networked mobile devices, the number of (indirectly) interconnected services in pervasive computing environments may grow without bound. The network contains a potentially enormous amount of context aware services that sense, gather and distribute context information ...

T;Context-Aware Exception Handling in Mobile Agent Systems? Handling erroneous conditions in context-aware mobile agent systems is challenging due to their intrinsic characteristics ...

T;Exploiting context awareness in information agents? mobile applications by incorporating explicit services to empower software agents with context-awareness ...

T;Debs 2010 context based computing tutorial? stributed topology: We have Context service on EPA level – The context ...

T;Improving Flood Warning times using Pervasive and Grid Computing? We are all aware of the increasing number of computing devices that we encounter on an everyday basis — not just the desktop and laptop personal computers but the mobile phones, PDAs (portable digital assistants), digital cameras and other electronic accessories that we carry on our person ...

T;Designing Distributed Applications using Mobile Agents? Motivation Mobile Agent technology Application domains MA frameworks overview MA based Structuring MA Framework Issues MA Application Case Studies Conclusion ...

F;The Foresight Academy of Technology? One popular mechanism for addressing this need is through the use of a drama ...

F;Self-Protected Mobile Agents? In this paper, we present a new solution for the implementation of ?exible protection mechanisms in the context of mobile agent systems, where security problems are currently a major issue ...

F;Method for packet data protocol context activation? A network requested PDP context activation (57) allows push applications (45) to unsolicited transmit push data to a mobile station (10). A gateway (30), GGSN (Gateway GPRS Support Node), typically initiates the network requested PDP context activation (57) ...

4.4 Analysis and Results

We had implemented the above mentioned datasets discussed in section and the interface and the information extracted is shown below:

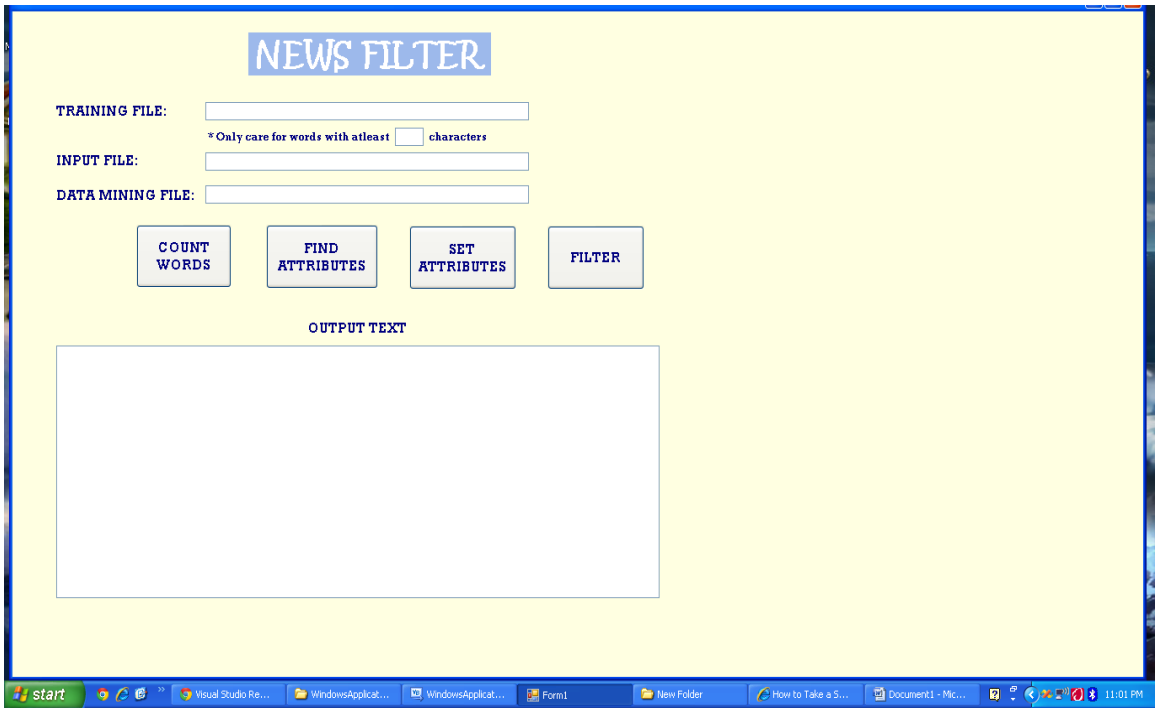


Figure 4.1 Main Interface

4.4.1 Word Count for the Data

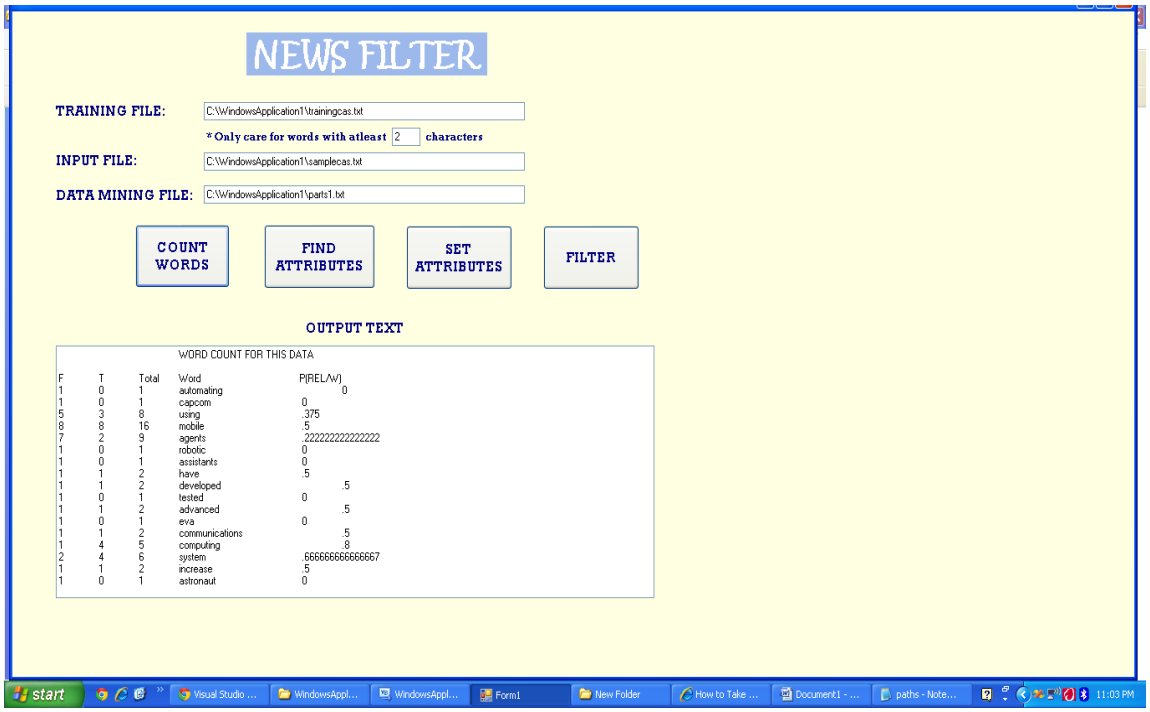


Figure 4.2 Word Count for the Data

F	T	Total	Word	P (REL/W)
1	0	1	automating	0
1	0	1	capcom	0
5	3	8	using	.375
8	8	16	mobile	.5
7	2	9	agents	.222
1	0	1	robotic	0
1	0	1	assistants	0
1	1	2	developed	.5
1	0	1	tested	0
1	1	2	advanced	.5
1	1	2	communications	.5
1	4	5	computing	.8
2	4	6	system	.666666666666667
1	1	2	increase	.5
1	0	1	astronaut	0
1	0	1	self-reliance	0

1	0	1	safety	0
1	0	1	roaming	0
3	0	3	security	0
1	0	1	guards	0
2	0	2	test	0
1	0	1	improve	0
2	4	6	paper	.666666666666667
1	0	1	discusses	0
1	0	1	design	0
1	1	2	implementation	.5
1	0	1	details	0
1	2	3	agent-based	.666666666666667
2	1	3	new	.333333333333333
3	2	5	network	.4
1	1	2	approach	.5
1	0	1	optimized	0
2	0	2	wireless	0
1	2	3	web	.666666666666667
1	0	1	browsing	0
3	3	6	use	.5
1	1	2	data	.5
1	0	1	comunications	0
2	0	2	increasing	0
1	0	1	rapidly	0
1	0	1	limited	0
1	0	1	bandwidth	0
1	0	1	reliability	0
8	5	13	agent	.384615384615385
7	10	17	ad	.588235294117647
2	1	3	hoc	.333333333333333
1	0	1	independent	0
1	0	1	piece	0
2	0	2	code	0
2	2	4	mobility	.5
1	0	1	autonomy	0
1	0	1	behavior	0
2	0	2	concepts	0
2	1	3	models	.333333333333333
1	0	1	tracy	0
1	0	1	toolkit	0
1	0	1	students	0
1	1	2	researchers	.5
1	0	1	introduction	0
1	0	1	possibilities	0
1	1	2	field	.5
1	0	1	overview	0

1	0	1	ongoing	0
1	2	3	research	.666666666666667
1	0	1	facilitating	0
1	3	4	information	.75
1	0	1	sharing	0
1	0	1	interests	0
1	0	1	rising	0
1	0	1	associated	0
1	0	1	employment	0
1	0	1	off-line	0
2	2	4	communication	.5
1	0	1	hosts	0
2	2	4	dynamic	.5
1	1	2	networks	.5
1	0	1	packet	0
1	0	1	radio	0
1	0	1	ad-hoc	0
1	0	1	node	0
1	0	1	router	0
1	0	1	secure	0
1	0	1	internet	0
1	1	2	applications	.5
2	3	5	based	.6
1	0	1	importance	0
1	0	1	motivated	0
1	0	1	exploration	0
2	0	2	execution	0
1	1	2	entities	.5
1	0	1	overcome	0
1	0	1	limits	0
1	0	1	client/server	0
2	1	3	model	.333333333333333
8	0	8	cas	0
1	0	1	community	0
1	0	1	central	0
1	0	1	authentication	0
2	6	8	service	.75
1	0	1	project	0
1	0	1	commonly	0
1	0	1	referred	0
2	0	2	computer	0
1	0	1	algebra	0
1	0	1	wikipedia	0
1	0	1	encyclopedia	0
1	0	1	software	0
1	0	1	program	0

1	0	1	facilitates	0
1	0	1	symbolic	0
1	0	1	mathematics	0
1	1	2	assist	.5
0	9	9	context	1
0	3	3	awareness	1
0	1	1	position	1
0	1	1	paradigm	1
0	2	2	important	1
1	2	3	technology	.666666666666667
0	1	1	enabling	1
0	2	2	pervasive	1
0	2	2	ubiquitous	1
0	1	1	context-based	1
0	4	4	addressing	1
2	1	3	concept	.333333333333333
0	1	1	large-scale	1
0	2	2	systems	1
0	2	2	notion	1
1	1	2	ability	.5
1	1	2	refer	.5
0	1	1	messages	1
0	1	1	collection	1
0	1	1	current	1
0	1	1	knowing	1
0	1	1	precise	1
0	1	1	identities	1
0	1	1	commerce	1
0	1	1	ieee	1
0	1	1	advances	1
0	1	1	e-commerce	1
0	1	1	resulted	1
0	1	1	significant	1
0	1	1	progress	1
0	1	1	strategies	1
0	2	2	requirements	1
0	2	2	development	1
0	1	1	major	1
0	1	1	dynamics	1
1	0	1	actuarial	0
1	0	1	society	0
1	0	1	knowledge	0
2	0	2	science	0
1	0	1	trade	0
1	0	1	promotion	0
1	0	1	management	0

1	0	1	optimization	0
1	0	1	retail	0
1	1	2	industry	.5
1	0	1	leader	0
1	0	1	solutions	0
1	0	1	consumer	0
1	0	1	university	0
1	0	1	aided	0
1	5	6	services	.833333333333333
1	0	1	cad	0
1	0	1	conversion	0
1	0	1	delivers	0
0	1	1	combat	1
0	1	1	operations	1
0	1	1	policy	1
0	1	1	constraints	1
1	0	1	find	0
1	0	1	list	0
1	0	1	logo	0
1	0	1	icai	0
15	8	23	ca	.347826086956522
1	0	1	shiksha	0
1	0	1	elearning	0
1	0	1	members	0
2	1	3	know	.333333333333333
1	0	1	rules	0
1	0	1	regulations	0
0	1	1	provisioning	1
0	1	1	sip	1
0	1	1	ambient	1
0	1	1	intelligence	1
1	3	4	application	.75
0	1	1	integrating	1
0	1	1	presents	1
0	1	1	impaired	1
0	1	1	people	1
0	1	1	explore	1
0	1	1	upper	1
0	1	1	level	1
0	1	1	bound	1
0	1	1	contains	1
0	1	1	potentially	1
0	1	1	enormous	1
0	1	1	amount	1
0	1	1	sense	1

4.4.2 Attributes for the Data

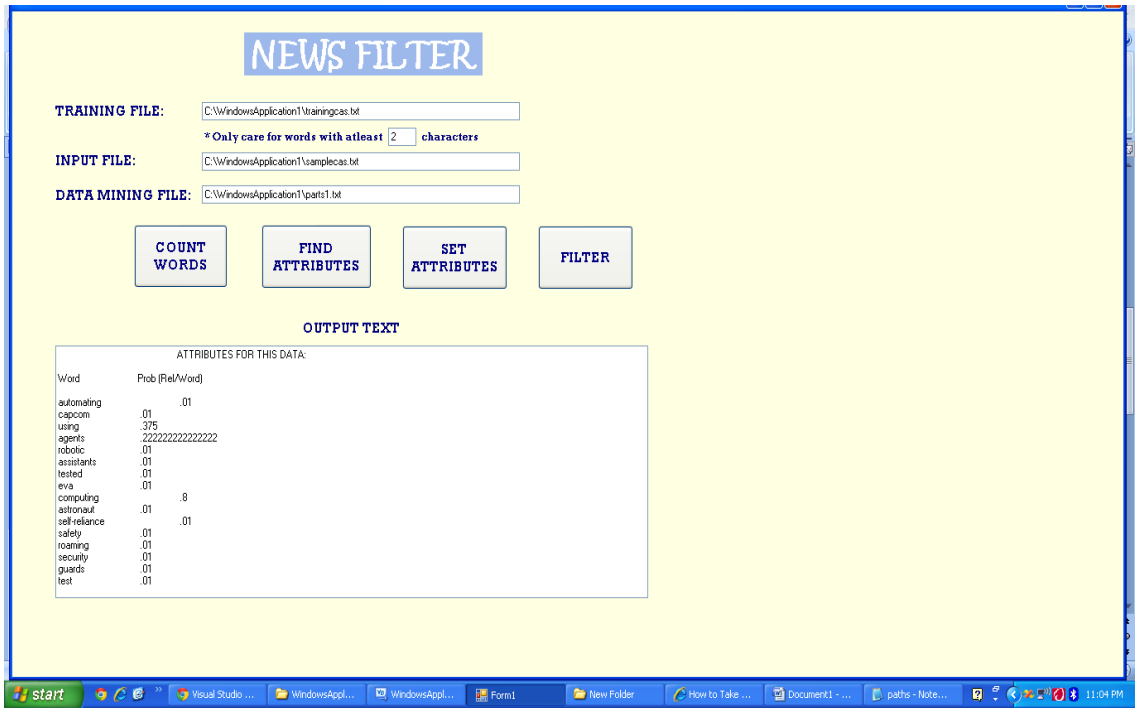


Figure 4.3 Attributes for the Data

Word	**Prob (Rel/Word)**
automating	.01
capcom	.01
using	.375
agents	.2222222
robotic	.01
assistants	.01
tested	.01
eva	.01
computing	.8
astronaut	.01
self-reliance	.01
safety	.01
roaming	.01
security	.01
guards	.01

Word	Prob (Rel/Word)
test	.01
improve	.01
discusses	.01
design	.01
details	.01
mast	.01
tool	.01
new	.333
network	.4
optimized	.01
wireless	.01
browsing	.01
comunications	.01
increasing	.01
rapidly	.01
limited	.01
bandwidth	.01
reliability	.01
agent	.3846
hoc	.333333333333333
simple	.01
independent	.01
piece	.01
code	.01
autonomy	.01
behavior	.01
basic	.01
concepts	.01
models	.333333333333333
tracy	.01
toolkit	.01
students	.01
book	.01
introduction	.01
possibilities	.01
overview	.01
ongoing	.01
facilitating	.01
sharing	.01
interests	.01
rising	.01

Word	**Prob (Rel/Word)**
benefits	.01
associated	.01
employment	.01
off-line	.01
hosts	.01
packet	.01
radio	.01
ad-hoc	.01
large-scale	.99
systems	.99
introduce	.99
notion	.99
send	.99
messages	.99
collection	.99
current	.99
knowing	.99
precise	.99
identities	.99
aware	.99
commerce	.99
ieee	.99
advances	.99
e-commerce	.99

4.4.3 Setting the Attributes for the Data

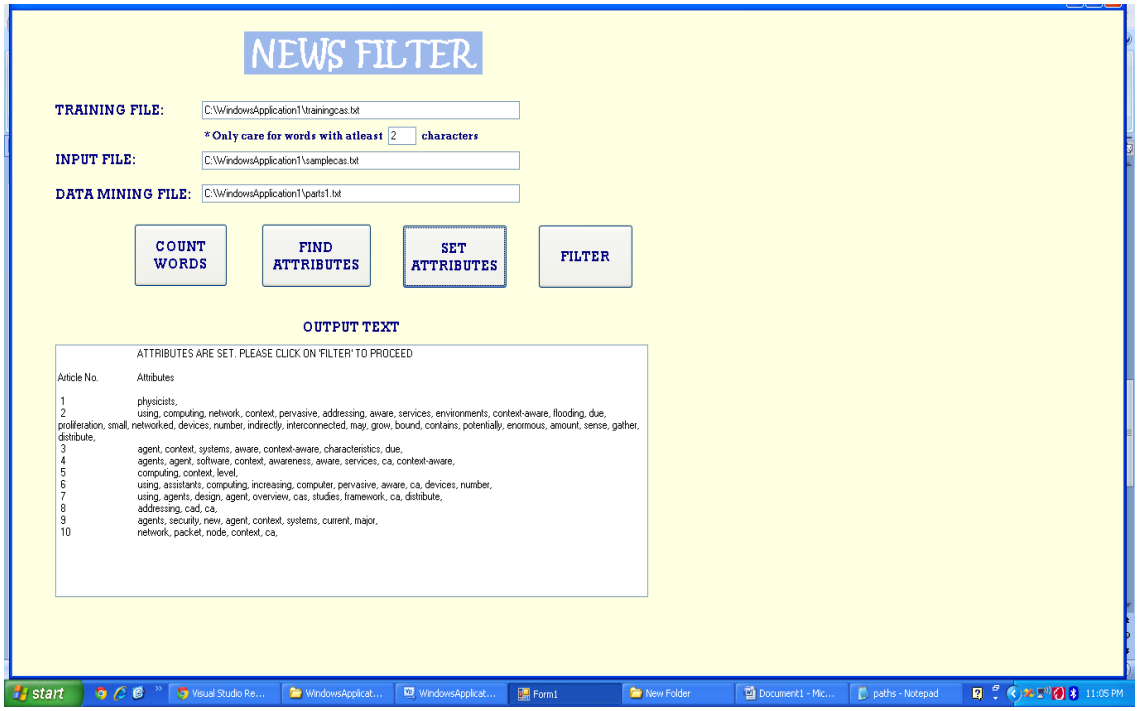


Figure 4.4 Setting the Attributes for the Data

Article no.	Attributes
1	physicists,
2	using, computing, network, context, pervasive, addressing, aware, services, environments, context-aware, flooding, due, proliferation, small, networked, devices, number, indirectly, interconnected, may, grow, bound, contains, potentially, enormous, amount, sense, gather, distribute,
3	agent, context, systems, aware, context-aware, characteristics, due,
4	agents, agent, software, context, awareness, aware, services, ca, context-aware
5	computing, context, level,
6	using, assistants, computing, increasing, computer, pervasive, aware, ca, devices, number,
7	using, agents, design, agent, overview, cas, studies, framework, ca, distribute,
8	addressing,
9	agents, security, new, agent, context, systems, current, major,
10	network, packet, node, context,

4.4.4 Filtered Articles

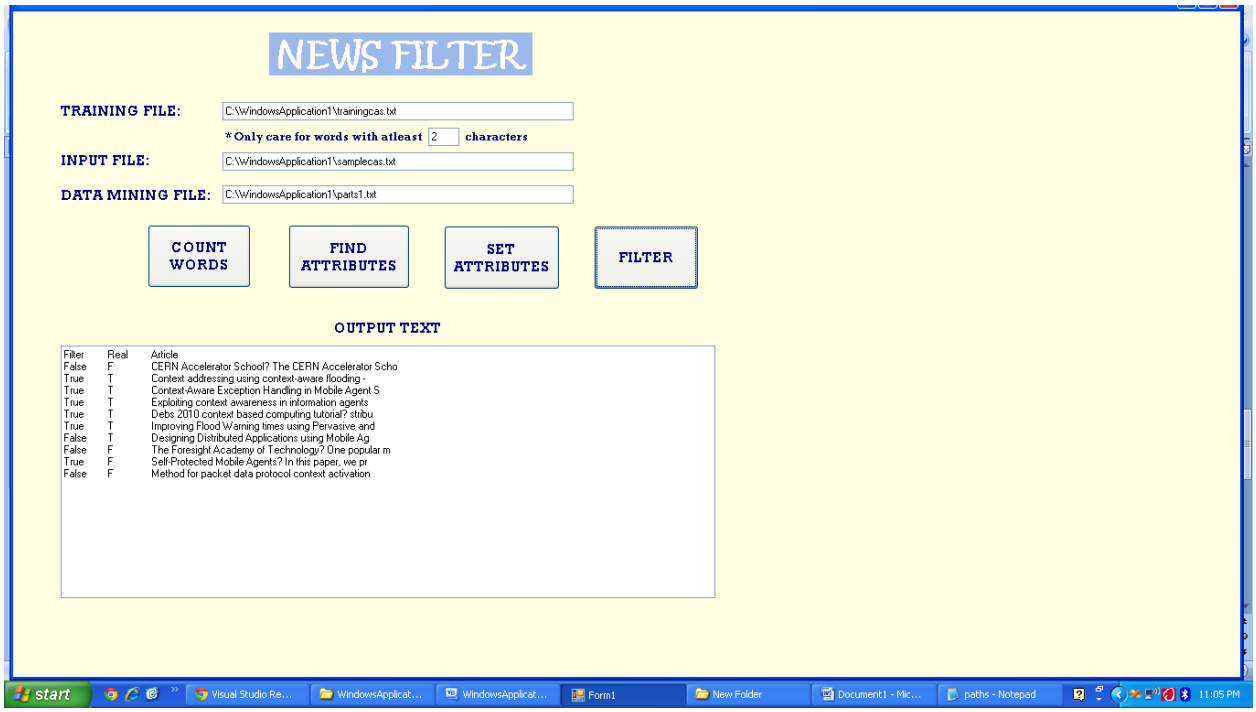


Figure 4.5 Final Filtering of the Articles

Filter	Real	Article
False	F	CERN Accelerator School? The CERN Accelerator Scho
True	T	Context addressing using context-aware flooding -
True	T	Context-Aware Exception Handling in Mobile Agent S
True	T	Exploiting context awareness in information agents
True	T	Debs 2010 context based computing tutorial? stribu
True	T	Improving Flood Warning times using Pervasive and
False	T	Designing Distributed Applications using Mobile Ag
False	F	The Foresight Academy of Technology? One popular m
True	T	Self-Protected Mobile Agents? In this paper, we pr
False	F	Method for packet data protocol context activation

4.5 Chapter Summary

In this chapter we had elaborated the implementation of the system based on the machine learning algorithm discussed in the previous chapters. We had also discussed all the datasets used in the system and how these datasets are used in the system at different steps. We had also discussed the analysis and result of the system and shows screenshots of every operation performed by the system to assist the new researchers.

Chapter 5

CONCLUSION AND FUTURE WORKS

5.1 Conclusion

The news filtering has been implemented successfully and it has shown that the filter is capable of filtering the news items that are relevant to the user, represented by letter T i.e. True and the news items which are irrelevant to the user , represented by letter F i.e. False.

To demonstrate the working of the project, Google news was used to fetch news for the search keyword. Any other search engine can also be used to do the same.

This application proves to be a good tool which can be used to search for any kind of results on any search engine, be it web search results, or news results or anything else. The Tool is accurate and has high operational speed. It takes less than 2 seconds to fetch, process, filter and display results for a search keyword.

The application is intelligent as it learns from experience. It takes in the user training and thus modifies its later results accordingly. Also the application is a personalized one as it is trained by the user and user can create profiles for each search keyword and save it in the Database and can later on search for the keyword again to see only filtered and processed results.

This News Filter thus is a useful and effective tool and solves the user's problems in searching.

5.2 Future Work

The model proposed and designed in this thesis is only being executed on the single keyword. The model can also be extended to search for news items which could be any group of keywords or it could be any phrase.

In this application we have first search the keyword and then apply the algorithm on the results that were fetched from the search engine. However, the application can be integrated with the Internet and fetches results directly from a search engine and automatically process them. We can also use a Database Management System to store and handle data for the application. Thus it could be a one stop application for the user, as he can type in the search keyword here, see the results, train it, and thus get filtered results, all at one place with an easy user friendly interface.

Bibliography

1. Xindong Wu, Gong-Qing Wu, Fei Xie, Zhu Zhu, and Xue-Gang, 2010. News Filtering and Summarization on the Web. *IEEE IntelligentSystems.*

2. Michael Chau , Hsinchun Chen, 2008. A machine learning approach to web page filtering using content and structure analysis. *Decision Support Systems - DSS, vol. 44, no. 2, pp. 482-494, 2008.*

3. Wei Peng, Juhua Chen and Haiping Zhou, 2009.An Implementation of ID3 --- Decision Tree Learning Algorithm,.

4. Sharad Verma, Nikita Jain, 2009. An Implementation of ID3 --- Decision Tree Learning Algorithm

5. Ethem Alpaydın, Introduction to Machine Learning, 2010, Second Edition, *The MIT Press Cambridge, Massachusetts London, England.*

6. Tom M. Mitchell, (1997). *Machine Learning,* Singapore, McGraw-Hill.

7. Daniela XHEMALI, Christopher J. HINDE and Roger G. STONE, 2009.Naïve Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages. *IJCSI International Journal of Computer Science Issues*.

8. Weka 3.6 - Data Mining with Open Source Machine Learning Software in java 1999-2003.

9. http://www2.cs.uregina.ca/~dbd/cs831/notes/ml/dtrees/4_dtrees2.html

10. H.Hamilton. E. Gurak, L. Findlater W. Olive**,** last modified 2011.Overview of Decision Trees.
http://www.cs.uregina.ca/~dbd/cs831/notes/ml/dtrees/4_dtrees1.html

11. YING HUANG, 2001. An Intelligent Adaptive News Filtering System.

12. David E. Goldberg and John H. Holland, Genetic Algorithms and Machine Learning.

13. Omar Jasso-Luna, Victor Sosa-Sosa and Ivan Lopez-Arevalo, 2008. An Approach to Building a Distributed ID3 Classifier. *International Symposium On Distributed Computing And Artificial Intelligence 2008 (Dcai 2008).*

14. Sally Jo Cunningham, James Littin and Ian H. Witten, *Applications Of Machine Learning In Information Retrieval.*

15. Lang, K. (1995) "NewsWeeder: Learning to filter Netnews." *Proceedings of the International Conference on Machine Learning*, (Tahoe City, California), pp. 331–339.

16. Carbonell, J. (Editor) (1990) *Machine learning: Paradigms and methods*. Bradford Books,MIT Press, Cambridge, Massachusetts.

17. Decision Tree Induction: Using Entropy for Attribute Selection, Principles of Data Mining.

18. Han Jiawei, Kamber Micheline, Second Edition, Data Mining:Concepts and Techniques.

19. N. J. Belkin and W. B. Croft. Information filtering and information retrieval: two sides of the same coin?, Volume 35 Issue 12, Dec. 1992. *Communications of the ACM - Special issue on information filtering.*

20. Sungjick Lee, Han-joon Kim, 2008.News Keyword Extraction for Topic Tracking. *Fourth International Conference on Networked Computing and Advanced Information Management*

Appendix A

A.1 Code of the System

```
Imports System.IO
Public Class Form1

    Private Sub TextBox3_TextChanged(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles TextBox3.TextChanged

    End Sub

    Private Sub Button1_Click(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles Count_words.Click

        Dim articles(,) As String = readInput(TextBox1.Text)
        Dim wordList(,) As String = findWords(articles, CInt(TextBox2.Text))

        setZeros(wordList)
        countWord(wordList, articles)

        writeOutput(wordList)
    End Sub
    Private Function readInput(ByVal inputFile) As String(,)
        'Open the file
        Dim orfReader As StreamReader
        orfReader = File.OpenText(inputFile)
        'Read throug the file and save to string array
        Dim articles(1, 0) As String
        Dim i As Integer = 0
        Do While orfReader.Peek > -1
            Dim line As String = orfReader.ReadLine
            ReDim Preserve articles(1, i)
            articles(0, i) = line.Substring(0, 1) ' Type (F = noisy, T = relevant)
            articles(1, i) = line.Substring(2) ' Article (includes title, name of newspaper, short
summary)
            i += 1
        Loop
        Return articles
    End Function
```

```
    Private Function findWords(ByVal articles(,) As String, ByVal leastChr As Integer)
As String(,)
        Dim wordList(3, 0) As String
        For i = 0 To articles.GetUpperBound(1) 'For each article
            Dim newsType As String = articles(0, i)
            Dim line As String = articles(1, i)
            Dim startChr As Integer = 0
            Dim afterChr As Integer = 1



            Dim startChr2 As Integer = 0
            Dim afterChr2 As Integer = 1

            Dim orfReader2 As StreamReader
            orfReader2 = File.OpenText(TextBox4.Text)
            Dim array2(0) As String
            Dim d As Integer = 0

            Do While orfReader2.Peek > -1
                Dim line2 As String = orfReader2.ReadLine
                line2.Trim()
                Dim drama As String = line2
                ReDim Preserve array2(d)
                array2(d) = drama
                'TextBox3.Text += array2(d)
                d = d + 1
            Loop
            Do
                If (line.Substring(afterChr, 1) = " " Or line.Substring(afterChr, 1) = "." Or
line.Substring(afterChr, 1) = ",") _
And afterChr - startChr >= leastChr Then 'Word found
                    Dim word As String = line.Substring(startChr, afterChr - startChr)
                    word = word.ToLower
                    word = word.Trim()
                    word = word.Trim(",", "?", "(", ")", "[", "]", ":", ";")
                    word = word.Trim(Chr(34)) 'Trim quotes
                    Dim flag As Integer = 0
                    For m = 0 To array2.GetUpperBound(0)
                        If word = array2(m) Then
                            flag = 1
                        End If
                        'TextBox3.Text += array2(m) + Str(value) + vbCrLf
                    Next
                    'TextBox3.Text += Str(flag)
                    If flag = 1 Then GoTo LK
```

```
                If word.Length >= leastChr Then 'If still long enough
                    If wordNew(word, wordList) Then 'Save if new word
                        If wordList(0, 0) <> "" Then 'Make space in the array when necessary
(for each word except the first one)
                            ReDim Preserve wordList(3, wordList.GetUpperBound(1) + 1)
                        End If
                        wordList(0, wordList.GetUpperBound(1)) = word
                    End If
                End If
LK:
                afterChr += 1
                startChr = afterChr 'Start looking for new word
            ElseIf line.Substring(afterChr, 1) = " " Or line.Substring(afterChr, 1) = "." Then
                'New word found, but too short
                afterChr += 1
                startChr = afterChr
            End If
            afterChr += 1


        Loop Until afterChr >= line.Length()
    Next
    Return wordList
End Function

Private Function findWords2(ByVal articles(,) As String, ByVal leastChr As Integer)
As String(,)
    Dim wordList(3, 0) As String
    For i = 0 To articles.GetUpperBound(1) 'For each article
        Dim newsType As String = articles(0, i)
        Dim line As String = articles(1, i)
        Dim startChr As Integer = 0
        Dim afterChr As Integer = 1
        Do
            If (line.Substring(afterChr, 1) = " " Or line.Substring(afterChr, 1) = ".") _
            And afterChr - startChr >= leastChr Then 'Word found
                Dim word As String = line.Substring(startChr, afterChr - startChr)
                word = word.ToLower
                word = word.Trim()
                word = word.Trim(",")
                word = word.Trim(Chr(34)) 'Trim quotes
                If word.Length >= leastChr Then 'If still long enough
                    If wordNew(word, wordList) Then 'Save if new word
                        wordList(0, wordList.GetUpperBound(1)) = word
                        ReDim Preserve wordList(3, wordList.GetUpperBound(1) + 1)
```

```
                End If
                'Update count
                updateCount(wordList, newsType, word)
            End If
            afterChr += 1
            startChr = afterChr 'Start looking for new word
        ElseIf line.Substring(afterChr, 1) = " " Or line.Substring(afterChr, 1) = "." Then
            'New word found, but too short
            afterChr += 1
            startChr = afterChr
        End If
        afterChr += 1
    Loop Until afterChr >= line.Length()
  Next
  Return wordList
End Function

Private Function wordNew(ByVal word As String, ByVal wordList(,) As String) As
Boolean
  'Returns true if the word is new. False otherwise.
  For i = 0 To wordList.GetUpperBound(1)
    If wordList(0, i) = word Then
      Return False
    End If
  Next
  Return True
End Function

Private Sub countWord(ByRef wordList(,) As String, ByVal articles(,) As String)
  For i = 0 To wordList.GetUpperBound(1) 'For all words
    For j = 0 To articles.GetUpperBound(1) 'Go through each article
      If wordExistsInLine(wordList(0, i), articles(1, j)) Then 'Update tallies
        If articles(0, j) = "F" Then
          wordList(1, i) = CInt(wordList(1, i)) + 1
        Else
          wordList(2, i) = CInt(wordList(2, i)) + 1
        End If
        wordList(3, i) = CInt(wordList(3, i)) + 1
      End If
    Next
  Next
End Sub
Private Function wordExistsInLine(ByVal word As String, ByVal line As String) As
Boolean
  Dim startChr As Integer = 0
  line = line.ToLower
```

```
        Do
            If line.Substring(startChr, word.Length()) = word Then
                Return True
            End If
            startChr += 1
        Loop Until startChr + word.Length() > line.Length()
        Return False
    End Function
    Private Sub updateCount(ByRef wordlist(,) As String, ByVal newsType As String,
ByVal word As String)
        For i = 0 To wordlist.GetUpperBound(1)
            If (wordlist(0, i) = word) Then
                wordlist(1, i) = CInt(wordlist(1, i)) + 1
                If newsType = "T" Then
                    wordlist(2, i) = CInt(wordlist(2, i)) + 1
                Else 'False
                    wordlist(3, i) = CInt(wordlist(3, i)) + 1
                End If
                Exit For
            End If
        Next
    End Sub
    Private Sub setZeros(ByRef wordList(,) As String)
        For i = 1 To 3
            For j = 0 To wordList.GetUpperBound(1)
                wordList(i, j) = "0"
            Next
        Next
    End Sub
    Private Sub writeOutput(ByVal wordlist(,) As String)
        TextBox3.Text = vbTab + vbTab + vbTab + "WORD COUNT FOR THIS DATA"
+ vbCrLf + vbCrLf
        TextBox3.Text += "F" + vbTab + "T" + vbTab + "Total" + vbTab + "Word" +
vbTab + vbTab + vbTab + "P(REL/W)" + vbTab + vbCrLf
        For i = 0 To wordlist.GetUpperBound(1)
            TextBox3.Text += wordlist(1, i) + vbTab + wordlist(2, i) + vbTab + wordlist(3, i)
+ vbTab + wordlist(0, i) + vbTab + vbTab + vbTab + Str(Val(wordlist(2, i)) /
Val(wordlist(3, i))) + vbTab + vbCrLf
        Next
    End Sub
    Private Sub writeAttributeOutput(ByVal attrcheck(,) As String, ByVal attr(,) As
String)
        'For i = 0 To attr.GetUpperBound(0)
        TextBox3.Text = vbTab + "ATTRIBUTES WERE SET. NOW PRESS 'FILTER'
TO START THE ALGORITHM"
        '  Next
```

```
        ' TextBox3.Text += vbCrLf
        ' For j = 0 To attrcheck.GetUpperBound(1)
        'For k = 0 To attr.GetUpperBound(0)
        'Dim l = (k + 1) Mod 7 'Becaue want print index 1, 2, ... 6, 0
        '  If attrcheck(l, j) = "" Then 'only True's have been set
        'attrcheck(l, j) = "F"
        '  End If
        ' TextBox3.Text += attrcheck(l, j) + vbTab
        '  Next
        '  TextBox3.Text += vbCrLf
        ' Next
    End Sub

    Private Function findattr(ByVal wordlist(,) As String)
        Dim attr(1, 0) As String
        Dim temp As Double
        Dim j As Integer = 0
        Dim b As Integer = 0
        Dim k As Integer = 0
        TextBox3.Text = vbTab + vbTab + vbTab + "ATTRIBUTES FOR THIS DATA:" +
vbCrLf + vbCrLf
        TextBox3.Text += "Word" + vbTab + vbTab + "Prob (Rel/Word)" + vbCrLf +
vbCrLf
        For i = 0 To wordlist.GetUpperBound(1)
            temp = Val(wordlist(2, i)) / Val(wordlist(3, i))
            If temp = 0 Then
                temp = 0.01
            End If
            If temp = 1 Then
                temp = 0.99
            End If
            If ((temp <= 1.0 And temp >= 0.8) Or (temp <= 0.45 And temp >= 0.0)) Then

                ReDim Preserve attr(1, b)
                attr(0, b) = wordlist(0, i)
                attr(1, b) = temp
                TextBox3.Text += attr(0, b) + vbTab + vbTab + Str(temp)
                TextBox3.Text += vbCrLf
                b = b + 1
            End If

        Next
        Return attr
    End Function
```

```
    Private Function setAttributes(ByVal articles2(,) As String, ByVal attr(,) As String) As
String(,)

        TextBox3.Text = vbTab + vbTab + "ATTRIBUTES ARE SET. PLEASE CLICK
ON 'FILTER' TO PROCEED" + vbCrLf + vbCrLf
        TextBox3.Text += "Article No." + vbTab + "Attributes"
        TextBox3.Text += vbCrLf + vbCrLf
        Dim attrcheck(attr.GetUpperBound(1), articles2.GetUpperBound(1)) As String

        For i = 0 To articles2.GetUpperBound(1)
            TextBox3.Text += Str(i + 1) + vbTab + vbTab
            For j = 0 To attr.GetUpperBound(1)
                If wordExistsInLine(attr(0, j), articles2(1, i)) Then
                    TextBox3.Text += attr(0, j) + ", "
                    attrcheck(j, i) = "T"
                    'Else
                    'TextBox3.Text += "N" + vbTab
                End If
            Next
            TextBox3.Text += vbCrLf
        Next
        Return attrcheck
    End Function

    Private Function filterNews(ByVal articles2(,) As String, ByVal attrcheck(,) As String,
ByVal attr(,) As String, ByVal threshold As Double) As Boolean()
        Dim relevant(articles2.GetUpperBound(1)) As Boolean
        Dim temp As Double = 0.0
        'Dim probRelevant(6) As Double
        'Assign values
        'probRelevant(1) = 0.01 'Name
        'probRelevant(2) = 0.05 'Entertainment
        'probRelevant(3) = 0.99 'Nano
        'probRelevant(4) = 0.85 'Metal
        'probRelevant(5) = 0.01 'Auto
        'probRelevant(6) = 0.9 'Mining
        'TextBox3.Text = ""
        TextBox3.Text = "Filter" + vbTab + "Real" + vbTab + "Article" + vbCrLf

        For i = 0 To articles2.GetUpperBound(1)
            Dim productProb As Double = 1
            Dim productInvProb As Double = 1

            For j = 0 To attrcheck.GetUpperBound(0)
                If attrcheck(j, i) = "T" Then
                    productProb *= attr(1, j)
```

```
                productInvProb *= (1 - attr(1, j))
            End If

        Next
        'TextBox3.Text += Str(productProb) + vbTab + Str(productInvProb) + vbTab
        temp = productProb / (productProb + productInvProb)
        'TextBox3.Text += Str(temp) + vbTab

        If temp > 0.5 Then
            TextBox3.Text += "True" & vbTab & articles2(0, i) + vbTab & articles2(1,
i).Substring(0, Math.Min(articles2(1, i).Length(), 50)) & vbCrLf
        Else
            TextBox3.Text += "False" & vbTab & articles2(0, i) + vbTab & articles2(1,
i).Substring(0, Math.Min(articles2(1, i).Length(), 50)) & vbCrLf
        End If
        'TextBox3.Text += Str(relevant(i)) + vbCrLf

    Next

    Return relevant
End Function

Private Sub Filter_Click(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles Filter.Click
    'Filter real news by assigning them a p(Relevant) and let go through if prob
>=threshold()
    Const threshold As Double = 0.5
    'Make array of news articles from input file
    Dim articles1(,) As String = readInput(TextBox1.Text)
    'Make array of attributes
    Dim articles2(,) As String = readInput(TextBox5.Text)

    Dim wordList(,) As String = findWords(articles1, CInt(TextBox2.Text))
    setZeros(wordList)
    countWord(wordList, articles1)
    Dim attr(,) As String = findattr(wordList)

    Dim attrcheck(,) As String = setAttributes(articles2, attr)
    'Set Relevant
    Dim relevant() As Boolean = filterNews(articles2, attrcheck, attr, threshold)
    'Print
    'printFilteredNews(relevant, articles2)
End Sub

Private Sub Button1_Click_1(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles Button1.Click
```

```
        'This program prints the attributes for each articles
        'The attributes are Name, Entertainment, Nano, Metal, Auto, Mining
        'Each attribute is by default False, but if a given word is found, it is changed to True
        'Make array of news articles from input file

        Dim articles1(,) As String = readInput(TextBox1.Text)
        Dim articles2(,) As String = readInput(TextBox5.Text)
        'Make array of attributes and classification

        Dim wordList(,) As String = findWords(articles1, CInt(TextBox2.Text))
        setZeros(wordList)
        countWord(wordList, articles1)
        Dim attr(,) As String = findattr(wordList)


        Dim attrcheck(,) As String = setAttributes(articles2, attr)
        'Print

        'Dim wordList(,) As String = findWords(articles, CInt(TextBox2.Text))
        'setZeros(wordList)
        'countWord(wordList, articles)
        'Dim attr(,) As String = findattr(wordList)

        'writeAttributeOutput(attributes, attr)
    End Sub

    Private Sub Button2_Click(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles Button2.Click

        Dim articles(,) As String = readInput(TextBox1.Text)

        Dim wordList(,) As String = findWords(articles, CInt(TextBox2.Text))
        setZeros(wordList)
        countWord(wordList, articles)
        Dim attr(,) As String = findattr(wordList)

    End Sub


    Private Sub Form1_Load(ByVal sender As System.Object, ByVal e As
System.EventArgs) Handles MyBase.Load

    End Sub
End Class
```

Appendix B

B.1 News Filter on Different Dataset

In this we are going to deal with another dataset to show that the news filter will work correctly for this dataset also.The dataset we are taking is for a specific search keyword ”Palladium” which is a rare metal. Let the user is intersted in all the news that infulence the prices of the metal Palladium. Apart form Palladium implying a metal, there are many other inference or application of this word like the names of several entertainment venues, brand names news paper, etc. So, we will apply the dataset to news filter and will observe the results.

B.1.1 Training Dataset for the Search Keyword “Palladium”

F;Today's Local Death Notices (updated throughout the day)? Palladium-Item - See complete obituary in Monday's Palladium-Item. McFarland, Earl, 75, Connersville, Ind., July 23, Showalter Blackwell Long Funeral Home, Connersville. ...

F;Legion zone tournament to conclude today? Herald Palladium (subscription) - By HP STAFF Heavy rain Friday night and early Saturday washed out all of Saturday's scheduled play in the American Legion zone baseball tournament at Eaton ...

F;Sports briefs? Herald Palladium (subscription) - By HP STAFF The Fruitbelt Officials Association has named two student-athletes as recipients of the group's inaugural scholarship awards for the 2009-10 ...

F;Be sure you have some kind of backup for your computer? Herald Palladium (subscription) - ... Long-Term Care at 800-654-2810 or check the Web site at www.areaagencyonaging.org. The Generations column appears each Sunday in The Herald-Palladium.

F;Democrat primary voters face choice in 79th? Herald Palladium (subscription) - William F. Ast III - Democratic candidates for the 79th District post in the Michigan House of Representatives are Mary E. Brown and ...

F;Fresh Start? Herald Palladium (subscription) - Scott Aiken - ST. JOSEPH - Judi plans to make changes when she gets out of jail, first by distancing herself from old acquaintances and ...

F;Lokey takes over SJHS orchestra? Herald Palladium (subscription) - Jeremy D. Bonfiglio - The 31-year-old viola player and music teacher who - for the time being - lives in Nyack, NY, was in town last ...

F;100 and counting? Herald Palladium (subscription) - Julie Swidwa - BENTON TOWNSHIP - The rain stopped in time for the celebration, but there was hardly a dry eye as Myron Stubbs was handed ...

F;Local sports calendar? Herald Palladium (subscription) - By HP STAFF COLOMA COMET FOOTBALL, Monday through Wednesday, east practice fields at Coloma High School. Grades 4 through 8 meet from 10 am to 1 pm Grades 9 ...

F;3 candidates seek Republican nomination in 79th District? Herald Palladium (subscription) - William F. Ast III - Republican candidates for the 79th District post in the Michigan House of Representatives are Bruce Gorenflo, ...

F;MEGADETH Plays 'Secret' Club Show In Montreal: Video Available ...? Blabbermouth.net - On March 31, 2010, at this incredibly special show at the legendary Hollywood Palladium, not far from where Dave Mustaine formed the band in 1983, ...

T;Researchers use nanotech to get power from sewage? Eetasia.com (subscription) - Oregon State University researchers found that nanoscale gold coating on standard graphite anodes boosted their efficiency more than standard palladium ...

T;Analysts bullish on zinc, lead in long term? The Australian - Robin Bromby - Magma Metals (MMW) reported more high-grade platinum, palladium, ... Last week BNP Paribas described palladium as the "prince of the precious metals". ...

F;Oberst & The Mystic Valley Band - The Sound Strike @ The Hollywood ...? Prefixmag - Andrew Martin - Just two days after Rick Ross appeared on Late Night with Jimmy Fallon to perform the original version of "BMF," Lupe Fiasco has released his own take on ...

F;Prefixmag Ima Castro Returns to the Recording Studio wutg Miguel Diaz for ...? Broadway World - Miguel joined the cast of The London Palladium production of The King And I. He played the role of the Interpreter and Kralahome, the prime minister. ...

T;Titanium-Jewelry.com Offers Free Ring Finger Sizing Kit? TechWhack (press release) - Enticing grooms and couples worldwide, the online retailer provides a multitude of styles and metals representing the best selection of palladium rings, ...

F;Crowd (un)control? Hometownlife.com - Jay M. Grossman - A fist fight between two girls in front of the Palladium. Police say the two girls, both teenagers, knew each other and were fighting over a text message ...

F;Producer Gary Gilbert gets red carpet treatment? Hometownlife.com - Prior to the screening at the Palladium, Gilbert's brother and sister-in-law ... The two also posed for photos on the red carpet at the Palladium and stayed ...

F;Luella Beach is honored on birthday; Dot Foods donates food to ... ?Palladium-Item - Rachel E. Sheeley writes "Everyday People" for the Palladium-Item. If you've got an interesting story, we'd love for you to share it with other area ...

F;Man cries NIMBY over possible plan for Napier offices? Herald Palladium (subscription) - Ralph Heibutzki - County Administrator Bill Wolf says concerns about a campus-type facility in Benton Twp. are premature By RALPH HEIBUTZKI - HP Correspondent BENTON TOWNSHIP ...

F;Atlantic City Food and Wine Festival returns to the resort? Press of Atlantic City - James Clark - Guy Fieri's Cheesesteak Battle, Caesars, Palladium Ballroom, 7 to 10 pm, ... Sweet & Stylish Hosted by Sandra Lee, Caesars, Palladium Ballroom, ...

F;Oregon Coast Events: Lincoln City August Calendar? OregonTravelDaily.com - Platinum/Palladium Printing 101 workshop with Rich Bergeman. Cost: $190, tuition and materials. 10 am – 4 pm. Sitka Center for Art & Ecology. ...

F;OregonTravelDaily.com Will Econ 101 stumpfour college seniors?? Washington Post - K.C. Summers - Grand Palladium White Sands Resort, Cancun: $704 per person through Apple Vacations. The 264-room oceanfront resort hasÂ seven freshwater swimming pools and ...

T;Nanotech coatings produce 20 times more electricity from sewage? Webnewswire.com (press release) - Coatings with palladium produced an increase, but not nearly as much. And the researchers believe nanoparticle coatings of iron – which would be a lot ...

T;Magma Metals hits high grade mineralized zones at Thunder Bay North? SteelGuru - ... Bridge Zone intersected several high grade mineralized zones at the Thunder Bay North platinum, palladium, copper, nickel project in northwest Ontario. ...

F;Boris Johnson: "Me? Prime Minister? I'll have a tilt at it'? Daily Mail - Jon Wilde - He has bounded onto the stage of the London Palladium and announced that he's just realised a lifelong ambition by emerging from the Judy Garland dressing ...

F;Daily Mail Today In Entertainment History July 25? WFMY News 2 - In 1980, Kiss introduced its new drummer, Eric Carr, at a concert at the New York Palladium. Carr replaced Peter Criss, who began a solo career. ...

T,Operational Excellence: The New Lever for Profitability and ...? Technology Evaluation Centers - The work integrates the strategy execution expertise of Palladium with the technology-enabled operational discipline of SAP. The study shows how operational ...

T;North American Palladium Ltd (PAL) Corporate Event Announcement Notice? Trading Markets (press release) - For full details on North American Palladium (PAL) PAL. North American Palladium (PAL) has Short Term PowerRatings at TradingMarkets. ...

F;Zack de la Rocha Dedicates "Killing in the Name" to Joe Arpaio? Phoenix New Times (blog) - Stephen Lemons - For a full report on Friday's Rage Against the Machine concert at the Hollywood Palladium, check out my write-up, here. It's so disappointing to see, Zack, ...

F;Doing the Right Thing? Collegenews.org - The awards were announced in a ceremony held July 19 at the Hollywood Palladium and broadcast live by VH1. Rembert was honored for co-founding Energize ...

T;Summary Box: Palladium, Platinum Prices Advance? ABC News - AP By AP GIVE 'EM A JUMP: Palladium and platinum, which are used in making catalytic converters, settled higher after Ford posted a strong profits and said ...

T;Palladium, platinum rise on improving auto sales? Washington Post - Sandy Shore - AP -- Palladium and platinum advanced Friday after strong earnings from Ford provided more evidence of a recovery in the auto industry. ...

T;North American Palladium to Host Second Quarter 2010 Results ...? MarketWatch (press release) - NAP is a Canadian precious metals company focused on the production of palladium and gold in mining-friendly jurisdictions. Lac des Iles, the Company's ...

T;Gold ends lower, stays flat for the week? MarketWatch - Claudia Assis - April H. Lee - Palladium for September delivery added $9.85, or 2.2%, to $466.75 an ounce. Platinum for October delivery rose $13.40, or 0.9%, to $1542.80 an ounce.

T;Commodity markets mostly climb, oil nears 80 dollars? MSN Malaysia News - On the London Platinum and Palladium Market, platinum climbed to 1541 dollars an ounce ... Palladium increased to 460 dollars an ounce from 456 dollars. ...

T;Palladium, platinum rise on improving auto sales? The Associated Press - Palladium and platinum are advancing on the latest sign of recovery in the auto industry. Prices for both metals, which are used in catalytic converters, ...

T;CORRECTED - PRECIOUS-Gold falls but holds ground after stress tests? Ninemsn - Among other precious metals, silver was at $18.08 an ounce against $18.07, while platinum was at $1539.50 an ounce versus $1521.10 and palladium at $465.50 ...

T;Dion's Friday ETF Winners and Losers? TheStreet.com - Don Dion - Two beneficiaries to a strong auto industry are platinum and palladium which are used extensively in the production of catalytic converters. ...

F;This Weekend: Miguel Atwood-Ferguson, Rage Against the Machine ...? LA Weekly (blog) - Kevin Bronson - Tonight's highlights: Rage Against the Machine, joined by Conor Oberst & the Mystic Valley Band, play the Palladium in a show protesting Arizona SB 1070. ...

T;Platinum, palladium to rise next year too? Commodity Online - LONDON (Commodity Online): The recent news that Euro zone economy is slowly gaining strength has helped platinum and palladium prices show signs of recovery ...

F;Local author reminisces about celebrity encounters? Winchester News Gazette - Bill Richmond - The book, due to be published around the first of October is based on Knight's experience as a district feature writer for the Richmond Palladium-Item. ...

T;Brussels Sprouts The Stress Test Report. Angelina Somewhat Insane ...? IBTimes - Jon Nadler - Palladium rose $4 to start at the $458.00 per ounce level and rhodium added $50 to achieve a bid quote of $2190.00 the ounce. Automaker Ford reported robust ...

T;Gold rises above $1200/oz ahead of stress tests? Economic Times - Among other precious metals, silver was at $18.13 an ounce against $18.07, while platinum was at $1538 an ounce versus $1521.10 and palladium at $458 ...

F;Sophie Evans to understudy as Dorothy in The Wizard of Oz? UK Theatre Tickets - The Wizard of Oz will open at the London Palladium on March 1st 2011 as the successor to Sister Act, which is now showing at the same venue. ...

F;No letdown for Silversun Pickups? The Columbian - Alan Sculley - The band tries to continuously evolve, even in a live setting, and seeks to insert new musical ideas into its existing ...

F;Rage Against The Machine To Play Protest Concert In Hometown? Thaindian.com - The band had held a press conference at Palladium, where they were joined by Chris Newman, who happens to be the lead counsel in the fight in opposition to ...

F;Cornell Adds Seven Recruits for the 2011 Season? LaxPower - ... 23 assists, 32 draw controls and 18 caused turnovers during her senior campaign and was named The Palladium-Times Female Athlete of the Year in 2010. ...

F;Friday's intriguing people? CNN (blog) - De La Rocha and other stars will perform Friday night at the Palladium in Los Angeles, California, under the banner "The Sound Strike Stop SB1070 Benefit ...

T;Melkior Intersects Significant Gold Mineralization in West Timmins? MarketWatch (press release) - Melkior also holds a 49% interest in the Delta Kenty nickel-copper-platinum-palladium deposit in Ungava and has several other gold properties in Ontario and ...

T;Top 5 Companies in the Precious Metals & Minerals Industry With ...? Comtex Smartrend - Chip Brian - Harry Winston Diamond (NYSE:HWD) follows with a beta of 1.8 and North American Palladium (AMEX:PAL) rounds out the top five with a beta of 1.7. ...

T;Platinum to gain as demand recovers? NewsDay - Palladium prices are seen averaging $472 an ounce this year, ... Palladium strongly outperformed other precious metals in the first quarter, rising 17,6% ...

T;North American Palladium Ltd. Makes Bullish Break; PAL, BBL, JRCC? Learning Markets - North American Palladium Ltd. (PAL) [Chart - Analysis - News] had a positive breakout as it jumped 6.51% in trading yesterday. PAL is considered a penny ...

F;Reporter wants to talk to local homeschoolers? Palladium-Item - The Palladium-Item would like to speak to anyone who recently left public school to be homeschooled. Reporter Brian Zimmerman would like to talk to parents ...

F;Longtime funeral director, Wayne County coroner remembered? Palladium-Item - Rachel E. Sheeley - In two separate Palladium-Item stories, Patterson talked about his work as a ... He told the Palladium-Item that the process of sifting through the rubble ...

F;Hanson's Ideal Husband stars with Bond at Vaudeville? London Theatre Guide - Previously he appeared in new musical Marguerite in the West End and The Sound Of Music at the London Palladium. Bond has been seen on the London stage in ...

T;Amur Minerals raises £1.2m for Kun-Manie copper-nickel project ...? Proactive Investors UK - Sergei Balashov - The GKZ (The Russian State Committee on Reserves) estimate is for 3960 tonnes of cobalt, 189400 ounces of platinum and 213800 ounces of palladium for the ...

T;Proactive Investors UK PRECIOUS METALS: Spot Gold Up, Trade Thin Ahead Of Stress Tests? Zawya - Rhiannon Hoyle - Spot silver was 0.2% higher at $18.13/oz, spot platinum was 1% higher at $1537.10/oz and spot palladium was 0.4% higher at $456.13/oz. ...

F;Ray Lowry's work off to New York, Tokyo and ... the Valley? Manchester Evening News - Simon Coyle - He superimposed the album title on a photograph of Paul Simonon smashing his base guitar on stage during a concert at the Palladium in New York.

T;Melkior Grants 1900000 Stock Options? AUTO-MOBI.info (press release) - Melkior also holds a 49% interest in the Delta Kenty nickel-copper-platinum-palladium deposit in Ungava and has several other gold properties in Ontario and ...

B.1.2 Test Dataset for the Search Keyword “Palladium”

F;Concert Notice: OMG! Adam Lambert's Glam Nation Tour comes to the ...? Dallas Voice But now, he's scheduled to play the Palladium Ballroom Sept. 7. Tickets go on sale Friday at 10 am The site lists the price at $39.50. ...

T;Gold firms as dip to 3-mth lows prompts buying? Reuters Africa Palladium XPD= bid at $484.50 an ounce against $465.93. For related news and prices, click on the codes in brackets: Spot gold/silver XAU= XAG= ...

T;old holds firm as low prices attract buyers? Business Spectator Palladium meanwhile rose to a one-month high at $US487.25 an ounce and was later at $US482.45 versus $US465.93, helped by the weaker US dollar and ...

T;FASTMARKETS AFTERNOON NEWS DIGEST? FXstreet.com The Forex Market Palladium struck a fresh one-month high around midday on Thursday, as solid risk appetite and a weaker euro helped push prices higher. ...

T;Aquarius Platinum May Shut Blue Ridge for 7 Months? BusinessWeek - Carli Lourens Platinum group metals include platinum, rhodium and palladium amongst others. A safety review was started at Blue Ridge after two deaths in the quarter ...

T;Go West, Young Man. Waaay West.? IBTimes - Jon Nadler Silver climbed 2 pennies to $17.49 but platinum added $9 (to reach $1546.00) and palladium rose $10 (to $478.00) per ounce. ...

T;Gold rises on Asian demand after price dip? Vancouver Sun - Jan Harvey Silver was at $17.56 an ounce against $17.44, platinum was at $1542 an ounce versus $1531.75 and palladium at $470.63 versus $465.93. ...

T;Stillwater Mining posts 18.2 per cent pgm output drop for Q2 2010? Platinum today The Montana-based firm revealed that it produced 112600 oz of platinum and palladium, compared to a figure of 137700 oz for the same period 12 months ...

T;Gold Gains in New York as Prices Near 3-Month Low Spur Demand? BusinessWeek - Nicholas Larkin - Kim Kyoungwha ... delivery added 0.7 percent to $1552.40 an ounce. Palladium for September delivery gained as much as 3.5 percent to a five-week high of $484.95 an ounce.

T;South Africa's Aquarius Platinum PGMs output up in June quarter? Platts Aquarius Platinum had attributable platinum group metals production of 110474 oz of 4E PGMs (platinum, palladium, rhodium and gold) in the quarter to June ...

F;EXPLORE THE MAGIC OF MEDAN? TravPR.com (press release) - It is possible to buy various designer goods from this mall, while other smaller malls include Medan Fair Plaza and Grand Palladium Hall are great places to ...

T;Gold eases on technical selling? Globe and Mail - Frank Tang - Jan Harvey - Platinum PL-FT was at $1592 an ounce against $1566.55, while palladium PA-FT was at $510 against $491, having hit its highest since mid-May earlier in the ...

T;Gold Futures Fluctuate as Euro's Rebound May Lessen Demand? BusinessWeek - Palladium futures for September delivery advanced $14.90, or 3 percent, to $514.90 an ounce. ---With assistance from Chanyaporn Chanjaroen in London. ...

F;Proos tops in money game? Herald Palladium (subscription) - Jim Dalgleish - John Proos is the runaway winner in Southwest Michigan's campaign money game on the eve of Tuesday's primary ...

F;Counseling helps unlock secrets of the psyche? Herald Palladium (subscription) - Glenn Chapman - Question: My husband and I are having a more or less friendly argument about the value of counseling. He thinks people ought to be able to solve their own ...

F;Question and Answer: Kelly Travis? Herald Palladium (subscription) - Shawn McGrath - She recently sat down with Herald-Palladium staff writer Shawn McGrath to talk about why she became a prosecutor and how she spends her time when she's not ...

F;Ben Sanders: No move was the right move for Tigers? Herald Palladium (subscription) - The Tigers can make their big moves in 2011. Doing so now wouldn't have saved 2010. Ben Sanders is a sports writer for The Herald-Palladium. ...

F;Sanders' baker's dozen? Herald Palladium (subscription) - Jim Dalgleish - ST. JOSEPH - Jan Leach said she's hard-pressed to think her mother ever had a better day than the one she had ...

F;SJ native offers dose of sanity for crazy times? Herald Palladium (subscription) - John Matuszak - As a human resources professional for General Motors, St. Joseph native Theresa Rich has been at the ...

T;Gold recovers; oil, copper surge? Economic Times - Platinum was at $1579.50 an ounce against $1566.55, while palladium was at ... Palladium reached a fresh 10-week high at $498 an ounce in earlier trade. ...

F;Fears, Worries, Yada Yada Yada: The New York Times Quails Before ...? Big Journalism (blog) - The right of the citizens to keep and bear arms has justly been considered, as the palladium of the liberties of a republic; since it offers a strong moral ...

F;K-SWISS Teams Up With Legendary HBO Cult Hero Kenny Powers for the ...? PR Newswire (press release) - K-Swiss also designs, develops and markets footwear under the Palladium brand, and owns the FORM Athletics brand. For more information about K-Swiss, ...

T;Gold Advances on Speculation Demand for Commodities Will Rise? BusinessWeek - Palladium futures for September delivery advanced $12.95, or 2.6 percent, to $512.95 an ounce. --With assistance from Chanyaporn Chanjaroen in London. ...

T;Stock to Watch: China Armco Metals -- August 2, 2010? SYS-CON Media (press release) (blog) - Stillwater Mining Company is the only US producer of palladium and platinum and is the largest primary producer of platinum group metals outside of South ...

T;PRECIOUS-Gold rises as euro strengthens, oil climbs Forexyard - Jan Harvey - Sue Thomas - Platinum was at $1587 an ounce against $1566.55, while palladium was at $501 ... Palladium reached a fresh 10-week high at $502 an ounce in earlier trade. ...

T;Stock to Watch: China Armco Metals -- August 2, 2010? Trading Markets (press release) - Stillwater Mining Company is the only US producer of palladium and platinum and is the largest primary producer of platinum group metals outside of South ...

T;Gold slips below $1180/oz as haven buying retreats? NASDAQ - Sue Thomas - Alison Birrane - Platinum was at $1579.50 an ounce against $1566.55, while palladium was at ... Palladium reached a fresh 10-week high at $498 an ounce in earlier trade. ...

T;Gold drops as investors shift to riskier markets? MarketWatch - Claudia Assis - In other metals, palladium and platinum were near multi-month highs. ... Palladium for September delivery added $12.20, or 2.4%, to $512.20 an ounce. ...

F;Home > Casino > Casino > Online Casino News > There's $100000 to ...? Bettingpro.com - One board is worth $35000 for regular players and the other, exclusively for the Palladium Lounge VIPs, that is worth more than $65000. ...

T;Crude Breaks Above 80 ahead of ISM? IBTimes - The benchmark contract for platinum rises for the 4th consecutive day to 1593, up +1% from last Friday's close while that for palladium continues staying ...

B. 1.3 Word Count for the Data

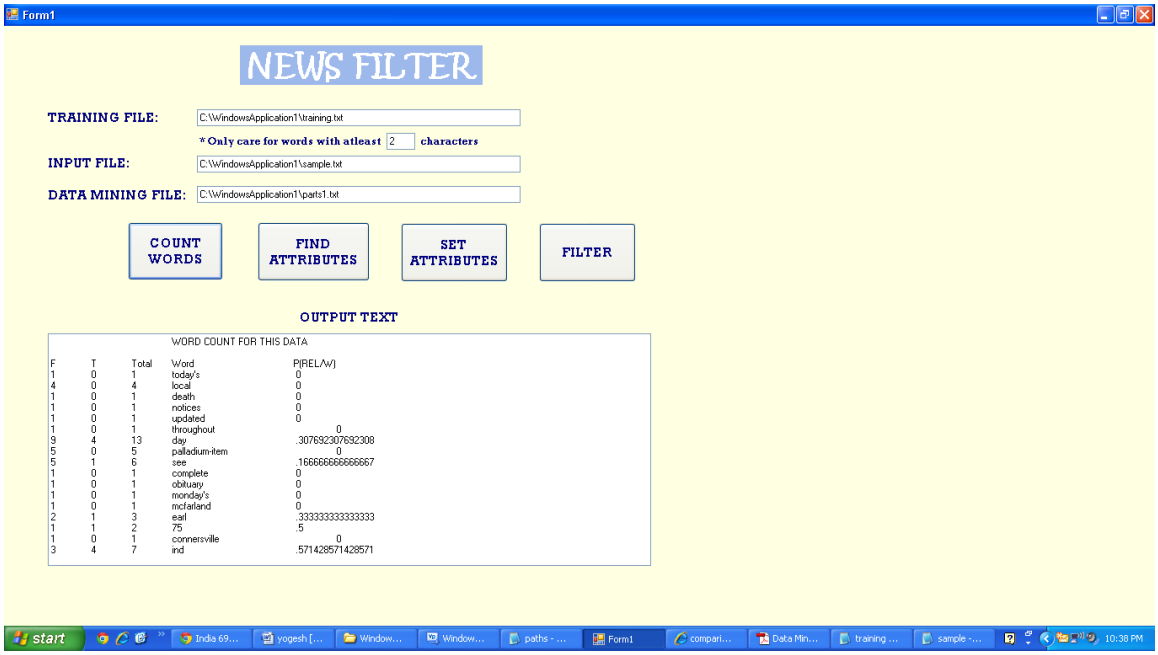


Figure B.1 Word Count for the Data

B1.4 Attributes for the Data

Figure B.2 Attributes for the Data

B1.5 Setting the Attributes for the Data

Figure B.3 Setting Attributes for the Data

B1.5 Final Filtering of the Data

Figure B.4 Filtering the Relevant and Irrelevant Data