Extraction of Pairs of Ailments Treatable by a Set of Medicinal Plants using an Apriori Algorithm

Victoria I. Yemi-Peters[1] and Bamidele A. Oluwade[2]
Department of Mathematical Sciences,
Faculty of Natural Science,
Kogi State University, Anyigba, Nigeria.
[1]*Email*: adeife74@gmail.com
[2]*Email*: deleoluwade@yahoo.com

ABSTRACT
In order to make healthcare delivery accessible to all especially the population in the rural areas, there is a need to embrace, once again the abandoned traditional medical care used by our forefathers. This is important at least for ailments that are common, rampant and intractable in curing in every environment in the African localities, since the modern healthcare facilities are not adequate, not affordable or sometimes not present in the Nigerian and African rural communities. The awareness and knowledge of medicinal properties of plants for preventive and curative purposes are inevitable. This paper explores the concept of association rule techniques, precisely the use of a classic algorithm of association of rules to list out the pairs of ailments associated to a set of medicinal plants from a developed data repository of African herbal information system. The paper presents a typical algorithm to extract pairs of ailments treatable by a set of medicinal plants using the concept of Apriori Algorithm. The programming and query languages used for the implementation are C#, ASP.Net and MS SQL.

Keywords: Ailments, Apriori Algorithm, Association rules, Frequent item set, Medicinal Plants

African Journal of Computing & ICT Reference Format:

Victoria I. Yemi-Peters and Bamidele A. Oluwade (2017),
Extraction of Pairs of Ailments Treatable by a Set of Medicinal
Plants using an Apriori Algorithm,
Afr. J. Comp. & ICT, Vol.10, Nos. 1& 2, pp. 37 - 50.

I. INTRODUCTION

From developments in contemporary nation, it is obvious that the use of herbal medicine cannot be ignored. There are numerous Chinese products like Tianshi, Forever Living Products, and Trevo being marketed in Nigeria today. All these are the Chinese traditional herbs being imported into Nigeria. Many Nigerians have subscribed to the use of these products and can testify that they are sometimes, more efficient and effective than the orthodox drugs. Indirectly, the economy of the Republic of China is being enhanced from Nigeria.

In order to make healthcare delivery accessible to all especially the population in the rural areas (being one of

the major goals of the World Health Organization), and to also control the spread of diseases, there is a need to embrace, once again, the abandoned traditional medical care used by our forefathers. This is at least necessary for ailments that are common, rampant and difficult to cure in every environment in our localities, since the modern healthcare facilities are not adequate or sometimes not present in the Nigerian and African rural communities. Nigeria as a developing nation, do not have a good enough ratio of hospital consultants for immediate attention in emergencies. Thus for medical facilities such as ours, rooted in traditional methods, there is an urgent need to avail the best in technology to maximize medical care and productivity [1].

The systems around the world are experiencing increased levels of chronic illness and escalating health care costs. Patients and healthcare providers alike are demanding that healthcare services be revitalized, with a stronger emphasis on individualized, person-centered care.

The awareness and knowledge of medicinal properties of plants in the African nations are passed practically or orally from one generation to the other. This, in the olden days, was practised by oracle priest, babalawo, boka and dibia as the case may be in the Nigeria major languages of Yoruba, Hausa and Ibo.. This study desires to use the application of information and communication technology to make available easy access to information and knowledge of the use of herbal plants to aid effectiveness and availability of healthcare delivery in Africa. The botanical names, common names, local names in the three major Nigerian languages, the diseases and ailments they can cure, are used as the template. An earlier work on screening of Nigerian plants for their medicinal plants without the use of ICT is [2]. A section of the repository database is used to illustrate the data mining concept, for individuals, pharmaceutical firms, researchers and professionals in the medical field to have an easy and quick access to information on medicinal plants that can be used to cure certain ailments.

The aim of this study is to extract pairs of ailments treatable by a set of medicinal plants using the concept of Apriori Algorithm. The objectives are as follows:

(i) To illustrate the use of Apriori Algorithm as a classification technique;

(ii) To identify or link the dataset (medicinal plants) that can cure a particular ailment or disease i.e. all the possible herbal plants which will constitute or be singly used as a medicine or curative mixture for the ailment;

(iii) To develop an application, using the Apriori algorithm, to extract pairs of ailments that frequently associate together, and are curable by a set of medicinal plants.

II. DATA MINING AND A PRIORI ALGORITHM

Data Mining is the practice of automatically searching large stores of data to discover patterns and trends that go beyond simple analysis. Knowledge is drawn from information which in turn is drawn from quality raw data. Information- based decision-making makes use of data and knowledge. The output of data mining is information pattern. Data mining has been defined as the practice of automatically searching large stores of data to discover patterns and trends that go, beyond simple analysis. Data mining uses sophisticated mathematical algorithms to segment the data and evaluate the probability of future events. Figure 1 shows the process involved in Data Mining [3].

In data mining, the important criteria for data are not the storage format, but its applicability to the problem to be solved. Data mining is a powerful tool that helps in the finding of patterns and relationships within a given set of data. It discovers the hidden information in the data but will not discover the value of the information [4]. The ailments are grouped in the classes of the communicable and non-communicable ailments. The concept of Apriori Algorithm is used to illustrate how an ailment is used to search out and classify the dataset of herbal plants that can be used or be a part of herbs for curative or medicinal mixture.

Apriori algorithm is the classic algorithm of association rules, which enumerates all of the frequent item sets. Apriori is designed to operate on databases containing transactions. The algorithm is used for association rule mining. Apriori is the best-known basic algorithm for mining frequent item sets in a set of transactions. It represents the candidate generation approach, which generates candidate (k+1) item sets based on frequent k-1 item sets [5, 6, 7, 8]. The Apriori algorithm has been used in many researches to implement the generation of association rule according to their minimum confidence threshold value in the area of mining frequent patterns, educational data pattern and also knowledge discovery for academic results [9, 10, 11].

The following are definitions of some of the other relevant terms which arise in this paper.

Item set: Item set is the collection of items in a database which is denoted by D = $\{x_1, x_2,\ldots, x_n\}$, Here 'n' is the number of items.

Candidate Item set: Candidate item sets are selected items that are to be considered for processing. Candidate item set are all the possible combination of item set; usually denoted by 'C_i' where 'i' indicates the i-th item set.

Transaction: Transaction is a database entry which contains collection of items. It is denoted by T ⊆ D. A transaction contains set of items T= $\{x_1, x_2,\ldots, x_n\}$.

Minimum Support: Minimum support is basically the condition to be satisfied by the given items so that further processing of that item can be completed. This condition helps in the removal of the items that are not frequent in the database. This is given in terms of percentage.

Frequent Item set: Frequent item set is the item set which satisfies the minimum support threshold value i.e. it is the commonly large item set.

Confidence: Confidence indicates the certainty of the rule. This argument lets us to count how often a transaction's item set pair with the left side of the implication with the right side. Any item set that do not satisfy the given condition can be discarded (using 60% for this illustration). Consider two items X and Y. To calculate confidence of X->Y the following formula is used, Conf(X->Y) = (number of transactions contained in both X & Y).

The step by step process of Apriori algorithm is stated below:

Step 0: Start

Step 1: Count the number of occurrence of each item in all transactions (frequency of occurrence).

Step 2: Select all items whose frequency of occurrence is greater or equal to the support S (Obtain single item that occur frequently).

Step 3: Make pairs for all elements selected in Step 2 and ensure only unique pairing.

Step 4: Count the number of occurrence of each pair in the transactions

Step 5: Repeat Step 2 for each pair (select pairs whose frequency is greater than support value i.e. the pairs of ailments having the same curative medicinal plants)

Step 6: To make 3 sets of ailments, find two pairs with the same medicinal plant, then combine without repeating any ailment.
Step 7: Stop.

The pseudocode of the above, as used in the implementation, is as follows:

III. IMPLEMENTATION

To achieve the implementation of the application for the extraction of the treatable ailments and the corresponding medicinal plants that can cure them, C#, object-oriented language, is used. The query languages used are ASP.Net and MS SQL.

Figures 3.1a and 3.1b below show screen shots for query codes for the implementation of the Apriori algorithm using the repository database of compiled 183 medicinal plants and the ailments they can cure. This shows the frequent occurring data items (ailments/diseases) in the repository database transactions that have met the condition of the minimum support threshold chosen, which is greater or equal 2 (≥ 2).

The query modules in Figure 3.2a and 3.2b implement the extraction of pairs or set of ailments that have frequent association together linked to the medicinal plants that can cure them.

IV. RESULT

The report of the application of Apriori algorithm for mining the repository database of the total 183 medicinal plants collated gave out the following classifications:

1. Set of Ailments/Diseases that have a frequent occurrence together in the repository database transactions. See Figure 4.1.
2. Figure 4.2 shows the output of the level of association of the frequent occurrences of ailments/diseases, confidence level and ratio of occurrence of ailments/diseases in the whole transaction of the database. This particularly helps to make the decisions on the combination of medicinal plants as constituents for drugs and herbal mixtures through the knowledge of the percentages of association of one ailment/disease to another.
3. Medicinal plants linked with the set of ailments/diseases that have frequent occurrence together in the repository database. See Figure 4.3.
   This shows the frequent occurring data items (ailments/diseases) in the repository database transactions that have met the condition of the minimum support threshold, which is greater or equal 2 (≥ 2). Table 4.1 is the summary of Figure 4.1 which gave an output of 157 transactions.

Table 4.2 is the summary of Figure 4.3 showing that there are five (5) instances where different medicinal plants can be used to cure four (4) sets of ailments/diseases that have frequent occurrence together. From the output, some medicinal plants appear more than once, which means they can cure other sets of ailments/diseases.

The implication of the above is further explained using a subset of the repository database structure shown in Table 4.3. In the table, some medicinal plants (on the second column) are selected as curatives to different ailments and preventives (on the 9th column). For the implementation, the number of occurrence of each ailment in each row is first read (Transaction), called frequency of occurrence. The occurrence of the following ailments is used: Arthritis, Cough, Asthma, Antimicrobial, Gonorrhea, Astringent. $T_{1,}\ldots,T_5$ represents the set of medicinal plants, as shown in Table 4.4.

From Table 4.5, given that level of confidence is met when an ailment has a frequency of a minimum of three (3) medicinal plants that can cure it, then, only the Ailment "Cough" met the golden rule of 60% occurrence that is, thrice. Hence, it shows that there occur three medicinal plants that can cure the ailment cough. These are Allium cepa (S/N 366), Allium sativum (S/N 367) and Allophylus africanus (S/N 550). It should be noted that the serial numbers are not arranged in any particular order, and so the last number in the column doesn't tally with the total number of medicinal plants.

V. CONCLUSION

Association rule mining, specifically the Aprori algorithm is a data mining technique that plays major role for extracting knowledge and updating of information. The implementation of this algorithm helps to save time and cost in searches and it also increases the efficiency of data mining in the search of medicinal plants to cure ailments and diseases. This

will assist researchers to make quick decisions on the production of curative herbal mixtures and drugs.

REFERENCES

[1] Agbachi, C. P. E. and Yemi-Peters, V.I (2016), Reachability in Telemedicine: Concepts, Design and Engineering, International Journal of Computer Trends and Technology, Vol. 35, No. 2, pp. 84-90.

[2] Fasola, Taiye R. (2000), Screening Nigerian Plants for their Medicinal Importance, Journal of Science Research (Journal of the Faculty of Science, University of Ibadan, Nigeria), Vol. 6, No. 1, pp. 51-57.

[3] Hongbo, D. (2010), Data Mining Techniques and Applications: An Introduction, Cengage Learning EMEA, United Kingdom.

[4] Bhardwaj, S., Chhikara, P., Vinayak, S., Pai, N., and Meena, K. (2015), Improved Apriori Algorithm for Association Rules, International Journal of Technical Research and Applications, Vol. 3, Issue 3 , pp. 238-240.

[5] Yabing, J. (2013), Research of an Improved Apriori Algorithm in Data Mining Association Rules, International Journal of Computer and Communication Engineering Vol. 2, No. 1, pp. 25-27.

[6] Goswami, D. N., Chaturvedi, Anshu; and Raghuvanshi, C. S. (2010), An Algorithm for Frequent Pattern Mining Based on Apriori, International Journal on Computer Science and Engineering, Vol. 2, No. 4, pp. 942-947.

[7] Shankar, S. and Purusothaman, T. (2009), Utility Sentient Frequent Itemset Mining and Association Rule Mining: A Literature Survey and Comparative Study, International Journal of Soft Computing Applications, Issue 4, pp. 81-95.

[8] Al-Maolegi, Mohammed and Arkok, Bassam (2014), An Improved Apriori Algorithm for Association Rules, International Journal on Natural Language Computing, Vol. 3, No. 1, pp. 21-29.

[9] Shukla, P. K. and Pandey, N. (2016), A Review on Mining Frequent Patterns and Association Rules Using Apriori Algorithm, International Journal of Innovative Research in Computer and Communication Engineering, Vol. 4, Issue 5, pp. 8647-8651.

[10] Oladipupo, O. O. and Oyelade, O. J. (2010) Knowledge Discovery from Students' Result Repository: Association Rule Mining Approach, International Journal of Computer Science & Security, Volume 4, Issue 2, pp. 199 - 207.

[11] Jha, J. and Ragha, L. (2013), Educational Data Mining using Improved Apriori Algorithm, International Journal of Information and Computation Technology, Vol. 3, No. 5, pp. 411-418

PSEUDOCODE

```
// W = ( ) ;
   //Array of Ailmemts
     Ailments = array (a1, a2,…, an);
     S = 0.6 * count (Transaction)
     Transaction=array (array1 (a1, a2,…,an),array2 (a1,
     a2…,an),…,arrayn (an (a1, a2, ..., an)
   // Array of frequency
   // frequency
   F = 0;
 For (int I = 0; I ≤ count (Ailments); i++)
 {  A = Ailment [i]
    For (int t =1; t ≤ count (Transaction); t++)
    {   if (in_array (A, Transaction))
            { F++; }
              If (F ≥ S) {
                 W ← W ∪ A;
                  }
    }  // obtaining the pair of
    Ailments
     int q = 0; F = 0; j = 1;
while (q ≤ count (W )
```

```
{
while (j ≤ count (W))
{
Pair = concatenate (W[q], W[j]);
For (K = 1; K ≤  count (Transaction); t++)
        If (in_array (pair, Transaction))
                {F++;
                  }
                If (F ≥ S) {
                cpair  ← cpair ∪ pair;
                }
                J++;
                }
                Q++; j = q + 1;
                }
```

Figure 0: Pseudocode

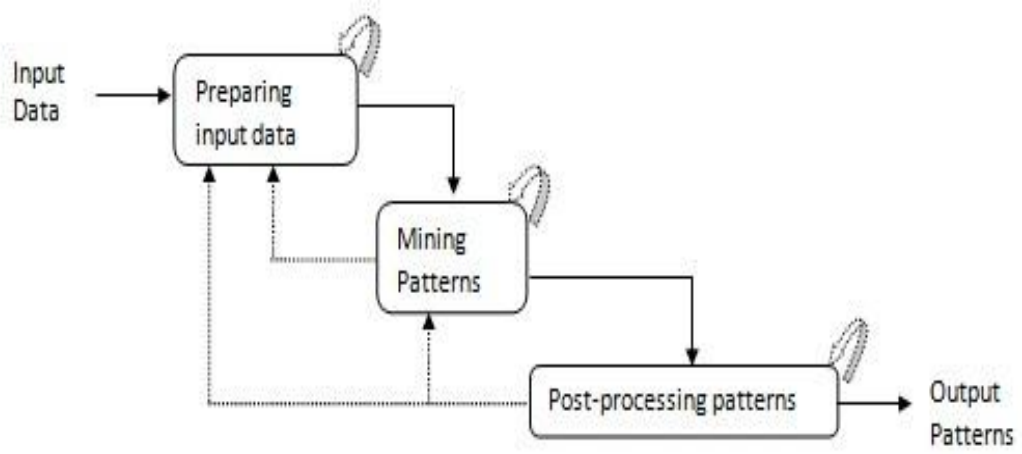


Figure.1: Data Mining Process (Source: [3])

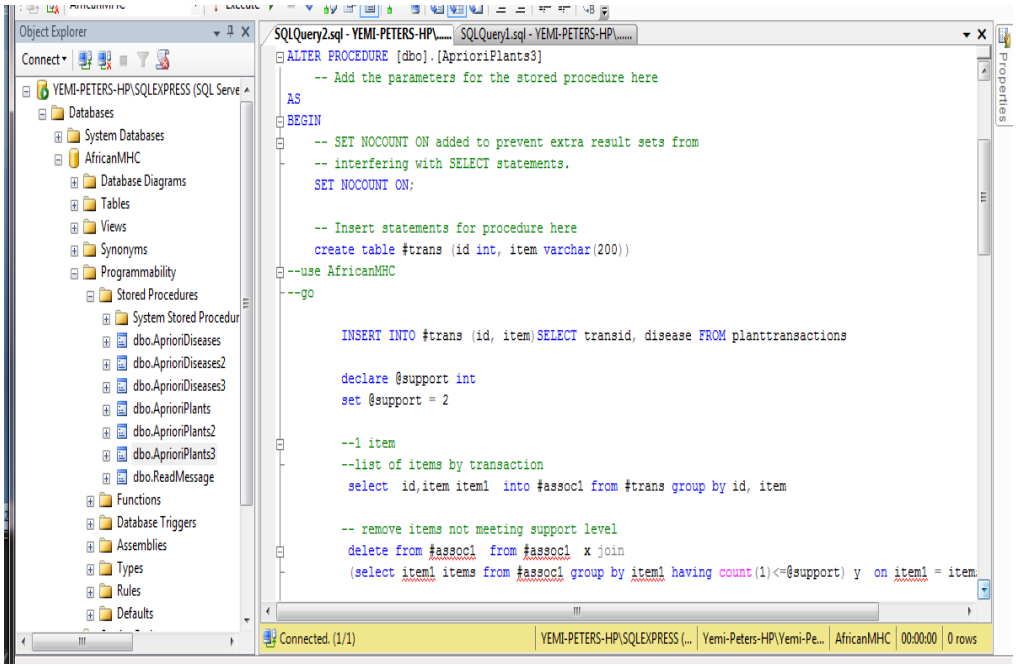


Figure 3.1a: Screenshot of Query code for implementing the frequent association of ailments.

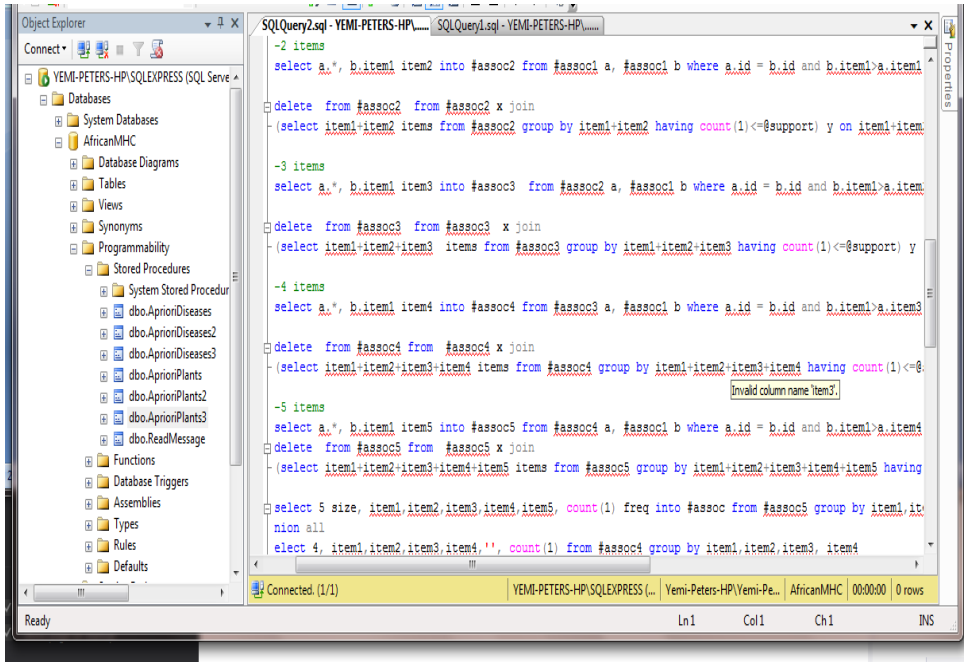


Figure 3.1b: Screenshot of Query code for implementing frequent association of ailments (continuation)

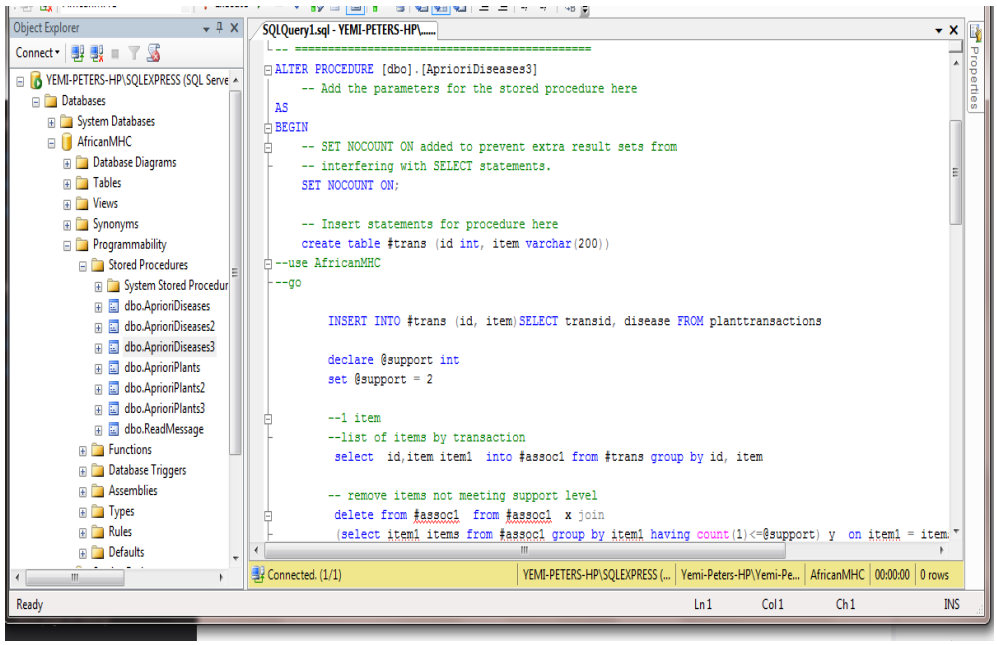


Figure 3.2a: Screenshot of query codes for implementing extracted ailments and the medicinal plants that can cure them.

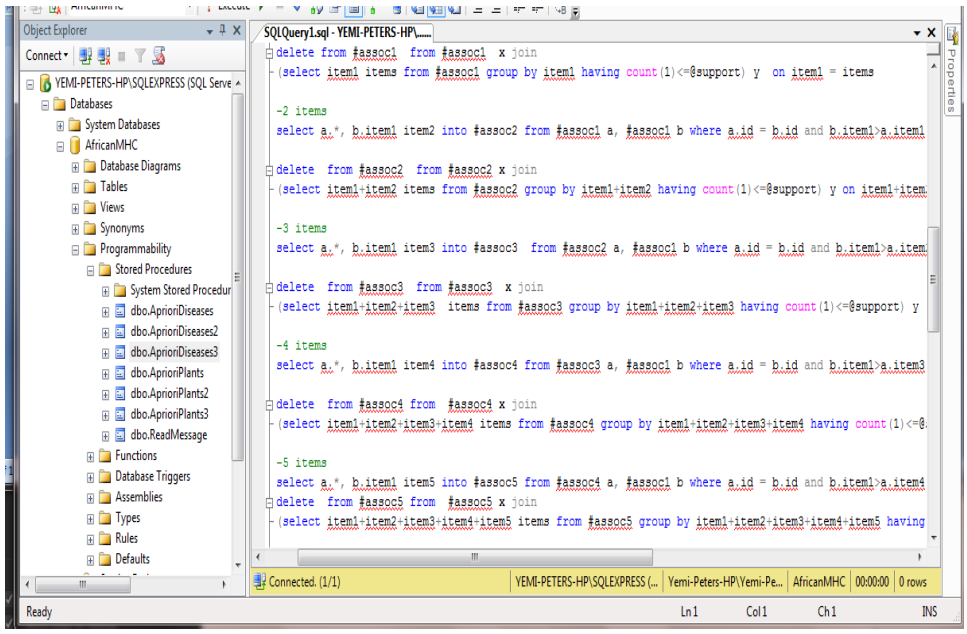


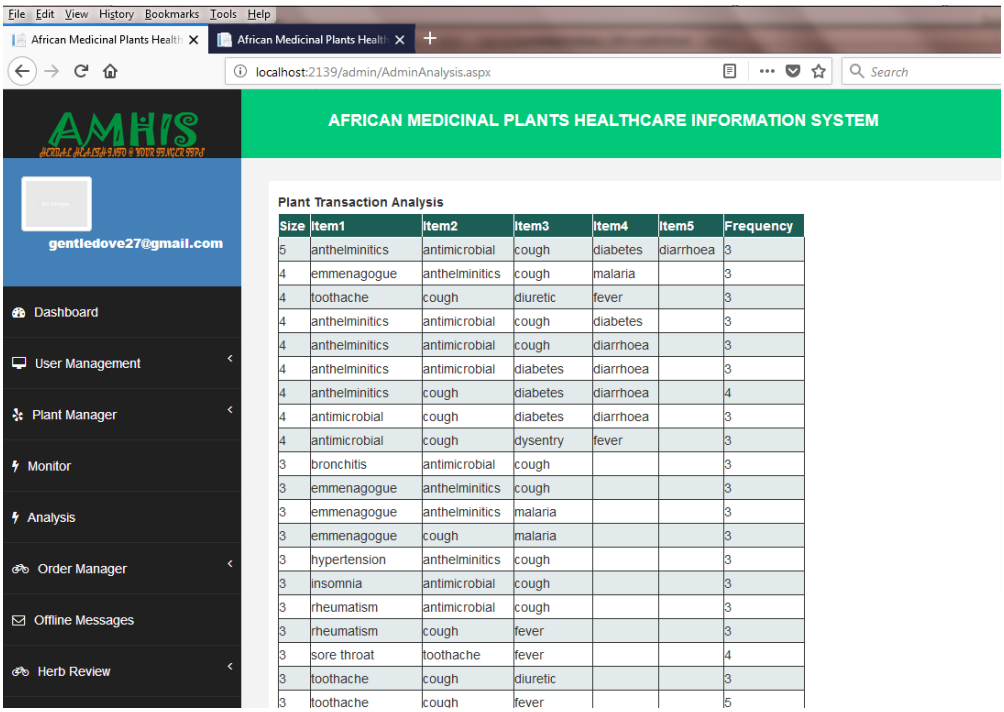
Figure 3.2b: Screenshot of query codes for implementing extracted ailments and the medicinal plants that can cure them.(continuation)

Number size of Ailments/Diseases that frequently occur together	Number of time(s) transaction appear
5	1
4	8
3	48
2	100

Table 4.1: summary of output of frequent occurring ailments/diseases

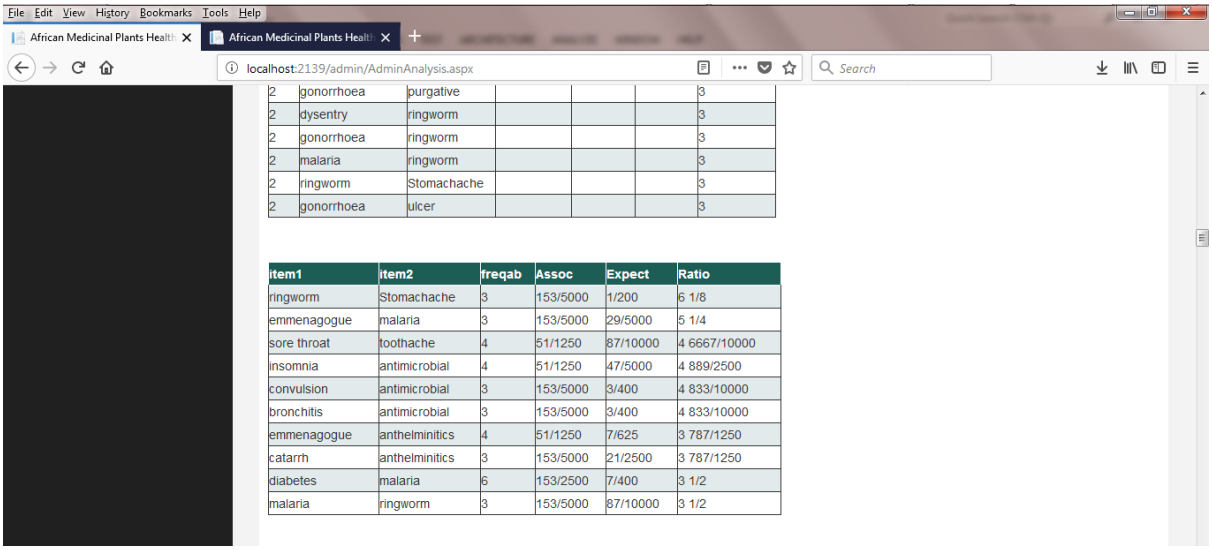
Number of Medicinal plant transactions	Size of combination of frequently occurring ailments/diseases
5	4
75	3
429	2

Table 4.2 Summary of frequent occurring diseases linked with medicinal plants



Size	Item1	Item2	Item3	Item4	Item5	Frequency
5	anthelminitics	antimicrobial	cough	diabetes	diarrhoea	3
4	emmenagogue	anthelminitics	cough	malaria		3
4	toothache	cough	diuretic	fever		3
4	anthelminitics	antimicrobial	cough	diabetes		3
4	anthelminitics	antimicrobial	cough	diarrhoea		3
4	anthelminitics	antimicrobial	diabetes	diarrhoea		3
4	anthelminitics	cough	diabetes	diarrhoea		4
4	antimicrobial	cough	diabetes	diarrhoea		3
4	antimicrobial	cough	dysentry	fever		3
3	bronchitis	antimicrobial	cough			3
3	emmenagogue	anthelminitics	cough			3
3	emmenagogue	anthelminitics	malaria			3
3	emmenagogue	cough	malaria			3
3	hypertension	anthelminitics	cough			3
3	insomnia	antimicrobial	cough			3
3	rheumatism	antimicrobial	cough			3
3	rheumatism	cough	fever			3
3	sore throat	toothache	fever			4
3	toothache	cough	diuretic			3
3	toothache	cough	fever			5

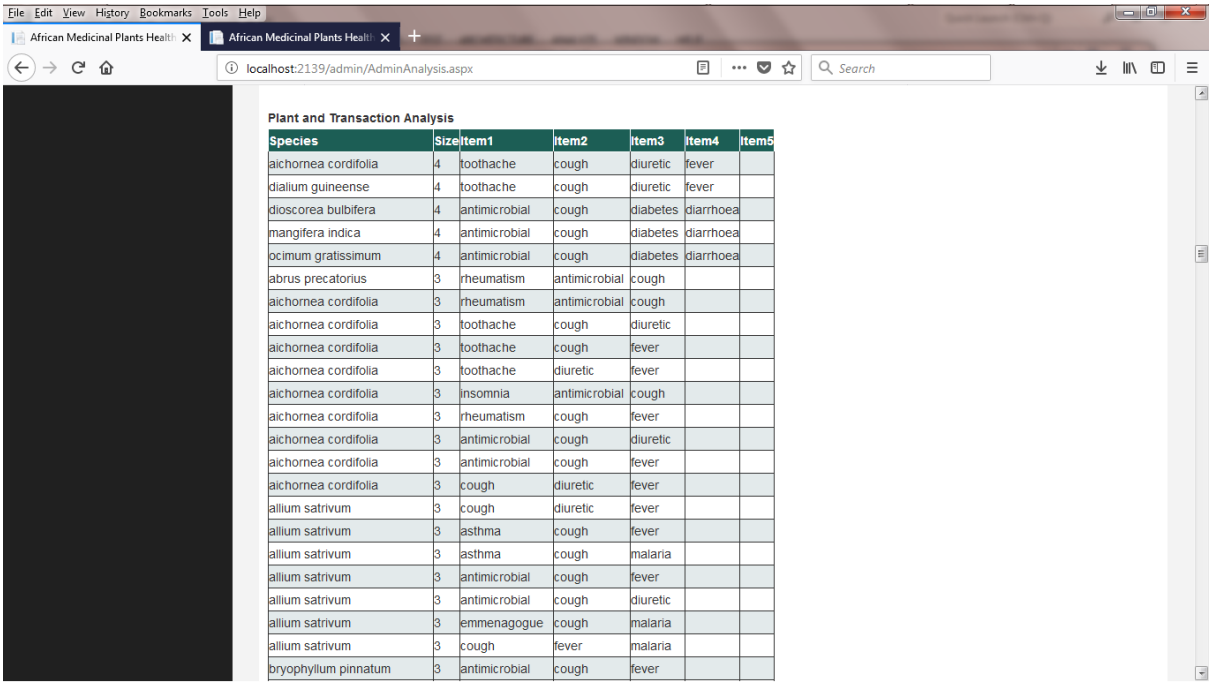
Figure 4.1: Interface showing the output of Apriori algorithm implementation (set of ailments/Diseases that have frequent occurrence).



2	gonorrhoea	purgative				3
2	dysentry	ringworm				3
2	gonorrhoea	ringworm				3
2	malaria	ringworm				3
2	ringworm	Stomachache				3
2	gonorrhoea	ulcer				3

item1	item2	freqab	Assoc	Expect	Ratio
ringworm	Stomachache	3	153/5000	1/200	6 1/8
emmenagogue	malaria	3	153/5000	29/5000	5 1/4
sore throat	toothache	4	51/1250	87/10000	4 6667/10000
insomnia	antimicrobial	4	51/1250	47/5000	4 889/2500
convulsion	antimicrobial	3	153/5000	3/400	4 833/10000
bronchitis	antimicrobial	3	153/5000	3/400	4 833/10000
emmenagogue	anthelminitics	4	51/1250	7/625	3 787/1250
catarrh	anthelminitics	3	153/5000	21/2500	3 787/1250
diabetes	malaria	6	153/2500	7/400	3 1/2
malaria	ringworm	3	153/5000	87/10000	3 1/2

Figure 4.2: Interface showing the output of Apriori algorithm implementation (Result of Association rule mining showing the level of Association and confidence Ratio level).



Species	Size	Item1	Item2	Item3	Item4	Item5
aichornea cordifolia	4	toothache	cough	diuretic	fever	
dialium guineense	4	toothache	cough	diuretic	fever	
dioscorea bulbifera	4	antimicrobial	cough	diabetes	diarrhoea	
mangifera indica	4	antimicrobial	cough	diabetes	diarrhoea	
ocimum gratissimum	4	antimicrobial	cough	diabetes	diarrhoea	
abrus precatorius	3	rheumatism	antimicrobial	cough		
aichornea cordifolia	3	rheumatism	antimicrobial	cough		
aichornea cordifolia	3	toothache	cough	diuretic		
aichornea cordifolia	3	toothache	cough	fever		
aichornea cordifolia	3	toothache	diuretic	fever		
aichornea cordifolia	3	insomnia	antimicrobial	cough		
aichornea cordifolia	3	rheumatism	cough	fever		
aichornea cordifolia	3	antimicrobial	cough	diuretic		
aichornea cordifolia	3	antimicrobial	cough	fever		
aichornea cordifolia	3	cough	diuretic	fever		
allium satrivum	3	cough	diuretic	fever		
allium satrivum	3	asthma	cough	fever		
allium satrivum	3	asthma	cough	malaria		
allium satrivum	3	antimicrobial	cough	fever		
allium satrivum	3	antimicrobial	cough	diuretic		
allium satrivum	3	emmenagogue	cough	malaria		
allium satrivum	3	cough	fever	malaria		
bryophyllum pinnatum	3	antimicrobial	cough	fever		

Figure 4.3: Interface showing the output of Apriori algorithm implementation (Medicinal Plants linked with the set of ailments/disease that frequently occur together).

Table 4.3: Repository Database Structure

S/N	FAMILY NAME	SPECIES NAME	YORUBA	HAUSA	IGBO	COMMON NAME	PARTS USED	MEDICINAL USE(S)
304	Leguminosae	Albizia zygia	Ayinre-weere	Madoobiiyar raafii	Nyie - avu	Okuro, atanza, siris	Bark	Arthritis, sprain
366	Liliaceae	Allium cepa	Alubosa	Albasaa, mai gudaji.	Yabaasi	Onion	Bulb, leaves	Cough, diuretic, anthelminitics, skin diseases, weak erection, anti-tumour, rubefacient, throat infection.
367	Liliaceae	Allium sativum	Ayo , Ayuu	Tafannuuwaa	Ayuu, Ayo-ishi	Garlic	Bulb	Fever, coughs, asthma, dilated bronchi, flatulence, anthelmintic, ringworm, antibiotic, diuretic, emmenagogue, antimicrobials, blood tonic, malaria.
550	Sapindacea	Allophylus africanus	Eekan-ehoro	Itaacen baka	Odu-oko akaito, Akato	African false currant	Leaves roots bark	Venereal diseases, cough, tuberculosis, urinary disorders, cancers, antimicrobials, sickle cell anaemia, emmenagogue, contraceptive, gonorrhoea, arthritis
17	Aniaranthace.a	Alternanthera sesilis	Reku-reku	Mai kai dubuu	Okbunzu nonu		Whole plant, leaves,	Astringent, antibactarial boil, headache, antidote to snake-bite.
18	Amaranthacea	Amaranthus spinosus	Tete-elegun-un, Dagunro	Namijin gaasayaa	Inine ogwu, Nnuno uku	Prickly amaranthus	Whole plant	Abdominal pain, ulcers, gonorrhoea, astringent, menorrhagia, diarrhea, dysentery, eczema

Table 4.4: Transactions of Ailments

Transaction ID	No. of Transaction
T1	Arthritis
T2	Cough
T3	Cough, Asthma, Antimicrobial
T4	Arthritis, Cough, Antimicrobial, Gonorrhea
T5	Astringent, Gonorrhea

Table 4.5: Occurrence of Ailment in Transactions

Ailment	No. of Occurrence in Transaction
Arthritis	2
Cough	3
Asthma	1
Antimicrobial	2
Astringent	1
Gonorrhea	2