ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 20, April 2016

Information Extraction Rules For S^2 Patient Clinical Data

L. Sathish Kumar[#1], Dr. A. Padmapriya[*2]

[#]*Ph.D. Research Scholar-Department of Computer Science and Engineering*
Alagappa University Karaikudi-630003. India.
[1]Lsathishkumarsva@yahoo.in

[*]*Associate Professor, Department of Computer Science and Engineering*
Alagappa University, Karaikudi-630003
[2]mailtopadhu@yahoo.co.in

***Abstract*— The research paper describes about information extraction rules (IER) for Semi-Structured (S^2) medical texts. The proposed work consists of two methods, the first one is pre-processing and information extraction rules generation. Second method is converting Semi-structure to structure data format for huge patient clinical data. The outcome of the method are prediction rules and disease patterns. These patterns are used to design many difficult templates.**

***Keywords*— Information Extraction rules, Semi structure, Patent Clinical Data.**

I. INTRODUCTION

Clinical data are rich source of information about the diseases, medical design and medical examination results [1]. Though Patient's records typically contain enormous information, it is not easy to extract information from it and also very challenging to analyse. Now a days it is quite common that all the health care institutions are having the practice of recording the medical examination descriptions, disease diagnosis, medicine prescribed and discharge summaries. The medical history of a patient generally includes medical reports, diagnoses and prescriptions given. These details are converted to digital form for further studies. This is also called as an electronic medical record. In order to preserve the details about a patient these electronic medical records must be secured from unauthorized access. But the electronic medical records of all health care institutions are not unique. This is primarily due to the variation in the way of storing the data. The form of the medical records is different form one health sector to another. So there is a need for converting the data to more standardized form such as databases. After that the information can be extracted with ease. ICD-10 [5] code format is very useful for extracting knowledge based medical data. With the help of the ICD-10 codes, we have the opportunity to communicate more clearly about the patient's condition. This process will enhance coverage, medical necessity and documentation. Once the diseases are classified in ICD-10, we can easily extract information from the electronic health care records using Rule-based systems. Rule based systems are used to extract information because the most common form of knowledge representation is If-then rules. The knowledge of the domain experts can be expressed and evaluated by rules. The information extracted from the rule based systems can be used in the process of decision making.

II. RELATION WITH OTHER WORK

In [2], the machine learning approach was introduced for identifying illness - handling relations in short medical texts. It mainly concentrated on entity recognition for diseases and treatment, relational discrimination by using hidden markov and maximum entropy replicas. There are three major methods used in extracting relations between entities, they are co-occurrences analysis, rule based approaches and stoical methods.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 20, April 2016

In [3] the context sensitive approach was introduced for medical information retrieval. This paper proposed new algorithms such learning procedure and retrieval procedure for classifying and choosing background in free-text medicinal descriptions. Sometimes the boundary of a sentence is not identified correctly.

The rule based [4] or similarly based systems require the users to input clear facts or attributes, such as symptoms of patients, to systems for determining the results. This is not easy for the user without medical background. Another disadvantage of this system is the requirement of clear definition for data attributes. Rule based systems are mostly used for decision making. It is mainly used for generating rules for patterning and identifying relations among the attributes.

III. IERS^2 BY FORWARD CHAINING

The architectural diagram of the proposed work is given below.

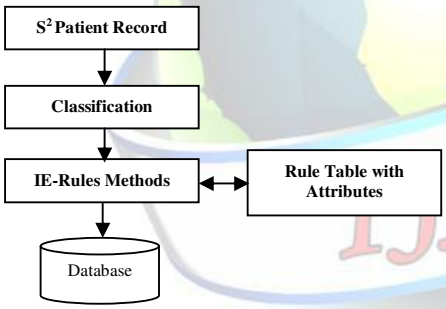

Fig 1. IERS^2 Architecture

The patient may get the test reports either as a hard copy or as an electronic health record. If it is in a manual for it has to be digitized otherwise the electronic health records are to be converted into more standardized form, i.e., tables or databases. Here the Information extraction rules are generated by two methods.

The algorithm for the first method is medical classification and ICD-10 code assignment to the patient's clinical laboratory reports is given below.

Procedure: Classification

Input: S^2 Patient Clinical Data
Output: Disease set Ranges

Begin

```
1   Input the medical Report values
2   Read medical Report Range Dataset
3   Read ICD-10 Database
4   j is value of the Test Range
5   for every possible range counts
6           if j < j1 then
7                   return range: low
8           else if j > j1 and j < j2 then
9                   return range: normal
10          else if j > j2 then
11                  return range: high
12          end if
13  next
14  compare range = ICD-10 code from disease
    table
15  find out the portions of the ICD-10 code for
    the report
16  return: disease and code
```

End: Classification

The laboratory test ranges and their values are stored in the classification table. The code for symptoms can be retrieved from ICD-10 database. All the information regarding the medical lab test such as lab name, values, ranges and symptoms are stored in the database. For illustration let us consider a hemogram report. The haemoglobin test value is 160, now check the rage of this value 160 that's returned range is low. Compare the range with ICD-10 classified database, and get all apt information based by haemoglobin value range.

Illustration:

The hemogram report will consists of the following attributes.

1. Hemoglobin

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 20, April 2016

2. PCV
3. RBC_count
4. WBC_count
5. ESR
6. Platelet_count
7. MCV
8. MCH
9. MCHC
10. Bleeding_time
11. Clotting_time

Classified Attributes and range:

Hemoglobin

Low Range: 12
High Range: 16
Low Range Disease: Aplastic anemia, Cirrhosis

High Range Disease: Chronic obstructive pulmonary disease, dehydration

PCV

Low Range: 37
High Range: 47
Low Range Disease: Anemia, down of erythrocytes in circulation
High Range Disease: Polycythemia Vera

RBC_count

Low Range: 4
High Range: 6
Low Range Disease: Anemia, less oxygen to tissues Tumours,
High Range Disease: Lung disease, low oxygen level

WBC_count

Low Range: 4000
High Range: 1000
Low Range Disease: Acute viral infection, cold, psittacosis
High Range Disease: Emotional stress, anesthesia, Convulsions

The algorithm for the second method rule based information extraction from electronic health record by backward chaining is given below.

Procedure: IERS^2

Input: Classified data and User Input
Output: Extract Rules and Disease

Begin

```
1   read prediction disease dataset
2   J is ICD-10 symptoms code
3   K is ICD-10 devices code
4   L is ICD-10 approach code
5   M is patient medical record dataset and D=0
6   if I = J then
7           return: symptoms and code
8   end if
9   if J = K then
10          return: devices and code
11  end if
12  if K = L then
13          return: approach and code
14  end if
15  for I to M (Count)
16          if I = M then
17                  D=D+1
18          end if
19  next
20  return: D
```

End: IERS^2

After the extraction of symptoms and ICD-10 code of selected value range using the defined rules. This algorithm will identify disease, devices and location codes for the matched symptoms code. The matched diseases are extracted and decide that disease is final disease. After that the health indicators can be efficiently extracted from the database using the rules generated by forward chaining.

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com
International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 3, Special Issue 20, April 2016

IV. EMPIRICAL FINDINGS

The empirical results for rule-based extraction method highly depend on the accuracy of medical texts. A human can easily identify, typographical errors such as the occurrence: 'I0 mm' in its place of '10 mm', are nearly incredible to fix by the algorithm lines. Since of this, medical documents that will suffer involuntary processing must be transcribed very prudently. Finally the information is signified by four attributes. In the table 1, four attributes has 100% precision but its low recall mainly that's why of missing and grammatical problems so the recall of 69.07% has the THERAPY_BEG attribute.

		Existing Algorithm				IERS2			
	Attribute	Cases	Prec.	Recall	F-Measure	Cases	Prec.	Recall	F-Measure
Various									
Reason of hospit.	REASON	83	98.73	93.98	96.3	84	99.61	90.98	97.3
Training of patient	EDUCATION	46	100	91.3	95.45	48	100	90.2	97
Beginning of insulin therapy	THERAPY_BEG	3	66.67	66.67	66.67	4	68.04	68.61	69.07
Therapy modification	THERAPY_ MODIF	23	23 100	91.3	95.45	36	94	95.3	98.45
Insulin dose modif.	DOSE_MODIF	10	90	90	90	10	90	90	90
Self monitoring	SELF_MONITORING	1	0	0	0	1	0	0	0
Diet correction	DIET_CORRECTION	1	100	100	100	1	100	100	100
Diet observing	DIET_OBSERVE	1	0	0	0	1	0	0	0

Table 1. Evaluation results for attributes found in 100 test diabetes documents

V. CONCLUSIONS

The research paper described the information extraction rules (IER) method for Semi-Structure (S^2) medical texts. In the proposed work classification and extraction information rules and converting Semi-structure to structure data format for huge patient clinical data. The proposed $IERS^2$ method will extract apt information and made standard templates. The method also achieves higher performance when compared with existing methods for information extraction rules.

REFERENCES

[1] AgnieszkaMykowiecka Małgorzata Marciniak Anna Kupść, " Information extraction from clinical data". Journal of Biomedical Informatics, Volume 42, Issue 5, Pages 923–936, October 2009.

[2] Razavan c. Bunuesu and Raymond j. Mooney, "Shortest path dependency kernel for relational extraction" human language technology conference on empirical method natural language processing. Pages 724-731, 10.3115/1220575.1220666. 2005.

[3] Mordechai averbunch, tom, karson "Context-sensitive medical information retrieval" health technology and informatics, 107(Pt 1):282-6. 2004.

[4] S.S abidi, and s. Manickam, "transforming xml-based electronic patients records for use in medical case based reasoning systems," medical infobahn for Europe.

[5] Measuring Diagnoses: ICD Code Accuracy, Kimberly J O'Malley, Karon F Cook, Matt D Price, Kimberly Raiford Wildes, John F Hurdle, and Carol M AshtonHealth Serv Res. Oct 2005; 40(5 Pt 2): 1620–1639. doi: 10.1111/j.1475-6773.2005.00444.x