**Ontology based Knowledge Healthcare System**

*Major Project Report*



*Submitted by:*

**Basit Halim(201600112)**

**Asifur Rahman(201600660)**

**Group Id: 87**

**PROJECT SUPERVISOR**

**Dr. Md. Tanwir Uddin Haider**

**Associate Professor, CSE**

**NIT PATNA**

**PROJECT REVIEWER**

**Mr. Dhruba Ningombam**

**Assistant Professor (SG), CSE**

**SMIT, Majitar**



**List of Contents**

|  |  |  |
| --- | --- | --- |
| **Sl No.** | **Title** | **Page no.** |
| I.  II.  III.  IV.  V.  VI.  VII.  VIII. | Abstract  Introduction  Literature Review  Problem Definition  Methodology  Conclusion and future works  Gantt Chart  References |  |

**ABSTRACT**

Ontologies have been considered as the backbone technology in most knowledge-based applications. Ontologies have become more common, their applicability has ranged from artificial intelligence areas such as knowledge representation and natural language processing to different fields such as information integration and retrieval systems, requirements analysis, and lately in semantic web applications. Ontology languages and their corresponding query languages play key roles for representing and processing information about the real world for the emerging Semantic Web. Knowledge based systems can be considered as a special type of database "that holds information representing the data of a particular domain. Rule based systems are one of knowledge based systems where the each rules can be expressed by If- Then statement. The if-part is the Left Hand Side (LHS), which is also called the antecedent. It consists of one or more of condition elements. The representation of the conditions may be categorised for simple problems, integer/real intervals or combination of these for more complex problems. The then-part -which is called the Right Hand Side (RHS), consequent or action-consists of number of actions. In Healthcare sector, one of the most valuable challenge is the one connected with the information extraction from raw data that implies the automatic detection of significant facts in unstructured texts and their transformation into structured documents, which are indexable and queryable exactly like databases.

Ontologies are used to capture knowledge about some domain of interest. An ontology describes the concepts in the domain and also the relationships that hold between those concepts. It has a richer set of operators - e.g. intersection, union and negation. It is based on a different logical model which makes it possible for concepts to be defined as well as described. OWL ontology consists of data and object Properties, Individuals and Classes, which roughly correspond to Protege frames, Instances, Slots and Classes. Complex concepts can therefore be built out of simpler concepts. Furthermore, the logical model allows the use of a reasoner which can check whether or not all of the statements and definitions in the ontology are mutually consistent, non redundant, logical and can also recognise which concepts fit under which definition. The reasoner can therefore help to maintain the hierarchy and structure correctly. This is particularly useful when dealing with cases where classes can have more than one parent.

**CHAPTER 1**

**INTRODUCTION**

**WEB ONTOLOGY LANGUAGE**

**What is ontology?**

Ontologies formalise the semantics of the domain explicitly by describing their elements; and thus, they consist of concepts that describe the internal features of the concepts, and the properties that describe the relationships between these concepts. Ontologies are based on a shared and consensual domain knowledge agreed by a community.

**Benefits of ontologies:**

1. The integration of heterogeneous data sources can benefit from the use of domain ontology to overcome semantic heterogeneities (Lacroix and Critchlow, 2003).
2. Ontology enables explicit and consensual knowledge to be shared and reused between human and software agents (Uschold and Jasper, 1999).
3. An ontology can be used to build knowledge bases - a knowledge base being an ontology with a set of instances (Noy and McGuinness, 2001). Also, ontologies can be used in deriving aspects of information systems at development or run time (Guarino, 1998). For example, ontology-based retrieval systems can assist users to browse and understand domain concepts, and therefore, formulate better specialised queries (Baker et. al, 1999).

**Types of ontology:**

Different kinds of ontologies exist that have been specified for different application domains thereby representing different types of knowledge. This section classifies ontologies along the following three dimensions: level of formality, level of generality, and primitive types

U7

**Methods and methodologies for developing ontologies**

1. Building ontology from scratch
2. Building ontologies from existing ontologies or from different data sources

The first activity in Grüninger and Fox methodology identifies the main scenarios that describe the purpose of the ontology with respect to the intended applications. Then, a set of competency questions are used to identify the scope of the ontology, thereby extracting the main concepts, properties, axioms of the underlying scope. After that, the elements of the ontology are expressed in first order logic.

The Uschold and King’s method proposes the following activities:

1. Identify the purpose of the ontology,
2. Build the ontology by capturing knowledge and identifying key concepts and properties in the domain, coding knowledge, and reusing other ontologies inside the current one,
3. Evaluate the ontology, and
4. Document the ontology.

The approach starts with a rough first pass at the ontology. This is followed by revising and refining the evolving ontology and filling in the details. Since building ontologies from scratch is not a simple task and is a time-consuming process, next we introduce the research work related to the second perspective, which studies the approaches for developing ontologies either from reusing existing ontologies or from reusing different data sources. For example, the developed ontology at Kactus (Bernaras et al, 1996) is built on the basis of an application knowledge base. In other words, the approach starts by building a knowledge base for an application. After that, when another knowledge base in a similar domain is needed, the first knowledge base can be generalised into an ontology. The output of repeating this process can lead to the development of an ontology that represents the consensual knowledge needed in all applications (Corcho et al, 2003). Furthermore, Maedche & Staab (2001) distinguished different approaches for developing ontologies from existing data sources based on the type of input. The input can be one of the following: (1) text where the ontology development is carried out by applying natural language analysis techniques to texts;

1. Dictionary where the relevant concepts and relations of an ontology is extracted from a machine readable dictionary;
2. Knowledge base; is used an existing source for building an ontology;
3. Semi-structured data is used for eliciting an ontology from sources which have any predefined structure;
4. Relational schema aims to extract relevant concepts, properties, relations from databases schema or relations.

**Ontology development engine architecture**

Fig. 1 illustrates the general framework to construct and develop an ontology based on the ruleset generated from previous discovery process. The architecture of the ontology development engine consists of the following phases.

1. Phase 1-Knowledge discovery and rules preparation: This phase is concerned with the extraction of patterns from the selected dataset over which a learning system, learning classifier system in particular, is applied. The generated rules are prepared in a suitable form to match the engine requirements.
2. Phase 2-Ontology development engine algorithm: This phase proposes a new algorithm to develop domain ontology from the generated ruleset. In this step, the ontology development engine considers a given domain ontology as a set of concepts used to describe a specific domain. The concepts are structured by two types of properties namely, subsumption and domain properties. The subsumption property represents the subtype relation in which one concept is more general than another whereas the domain property represents the relationships between domain concepts.

FIG 1: GENERAL FRAMEWORK [5]

6

**CHAPTER 2**

**LITERATURE REVIEW**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **S.No** | **Author** | **P a p e r** | **&** | **Findings** | | **Research Gap** |
| **.** |  | **p u b l i c a t i o n** | |  |  |  |
|  |  | **details** |  |  |  |  |
|  |  |  | |  |  |  |
| 1. | Flora Amato [1] | Building Ontology | | · | A new | The basics of |
|  |  | from knowledge |  |  | approach to | ontological |
|  |  | base Systems |  |  | develop certain | application and |
|  |  |  |  |  | domain | approach are |
|  |  |  |  |  | ontology. | learned through |
|  |  |  |  | · | The proposed | this p a p e r. Wo r |
|  |  |  |  |  | approach has | k i n g o n protege, |
|  |  |  |  |  | integrated | its tools and |
|  |  |  |  |  | different data | techniques were |
|  |  |  |  |  | mining | learned through |
|  |  |  |  |  | techniques to | this paper. |
|  |  |  |  |  | assist in |  |
|  |  |  |  |  | developing a |  |
|  |  |  |  |  | set of |  |
|  |  |  |  |  | representative |  |
|  |  |  |  |  | consensual |  |
|  |  |  |  |  | concepts of the |  |
|  |  |  |  |  | underlying |  |
|  |  |  |  |  | domain. |  |
|  |  |  |  |  |  |  |

7

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2. | Flora Amato, | A knowledge |  | · | A knowledge | | | The basic outlook | | | |
|  | Giovanni Cozzolino, | Based C o l l a b o |  |  | B a s e d | |  | o f | h o w | | a |
|  | Alessandro Maisto, | r a t i v e |  |  | Collaborative | | | knowledge | | based | |
|  | Antonino Mazzeo, | framework for e- |  |  | framework | | that | health care system | | | |
|  | Vincenzo Moscato, | health And |  |  | consists | in | a | were learned from | | | |
|  | Serena Pelosi, | Dipartimento di |  |  | doublefaced | | | this paper. Also the | | | |
|  | Antonio Picariello, | Ingegneria |  |  | system | aiming | | rough | idea | of | the |
|  | Sara Romano, Carlo | Elettrica e d e l l e |  |  | to s u p p o r t c | | | c l a s s e s | | a n d | |
|  | Sansone | T e c n o l o g i e |  |  | l i n i c a l | | | properties | | to be | |
|  |  | dell’Informazione. |  |  | processes in a | | | u s e d |  | w e r e | |
|  |  |  |  |  | more e f f e c t i | | | extracted. | |  |  |
|  |  |  |  |  | v e w a y | |  |  |  |  |  |
|  |  |  |  |  | (diagnostics | | |  |  |  |  |
|  |  |  |  |  | and therapy) | | |  |  |  |  |
|  |  |  |  |  | | | |  | | | |
| 3 | Dr.Sunita Abburu & | Ontology Driven | · | Ontology centric | | | | Advocates the use | | | |
|  | Suresh Babu Golla | Knowledge Based |  | approach provides | | | | of ontologies in | | | |
|  |  | Healthcare System |  | a |  |  | n | the project and not | | | |
|  |  |  |  | innovative ,effectiv | | | | a n y |  | o t h e r | |
|  |  |  |  | e | and efficient | | | t e c h n i q u e | | | a s |
|  |  |  |  | means of capturing | | | | ontology centric | | | |
|  |  |  |  | and organizing | | | | approach proves to | | | |
|  |  |  |  | k n o w l e d g e t o | | | | b e t h e m o s t | | | |
|  |  |  |  | r e p r e s e n t | | | t h e | efficient | |  |  |
|  |  |  |  | medical health care | | | |  |  |  |  |
|  |  |  |  | domain area for | | | |  |  |  |  |
|  |  |  |  | h e a l t h | | c a r e | |  |  |  |  |
|  |  |  |  | m a n a g e m e n t | | | |  |  |  |  |
|  |  |  |  | system | |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |  |  |

**CHAPTER 3**

**PROBLEM DEFINITION**

It has long been a desire in the field of research to develop a system which is able to map/predict the repercussions of the diagnosis available. Doctors very commonly have expressed an urge to be provided with such systems which eases the process of diagnosing the patient’s reports and mapping it with treatment. Systems currently available are mostly inaccurate and do not map the test data to diseases. They mainly focus upon treatment. There are many instances where patients did not have complete information and knowledge about their health condition. The dependency upon domain experts and doctors is high.

**CHAPTER 4**

**METHODOLOGY**

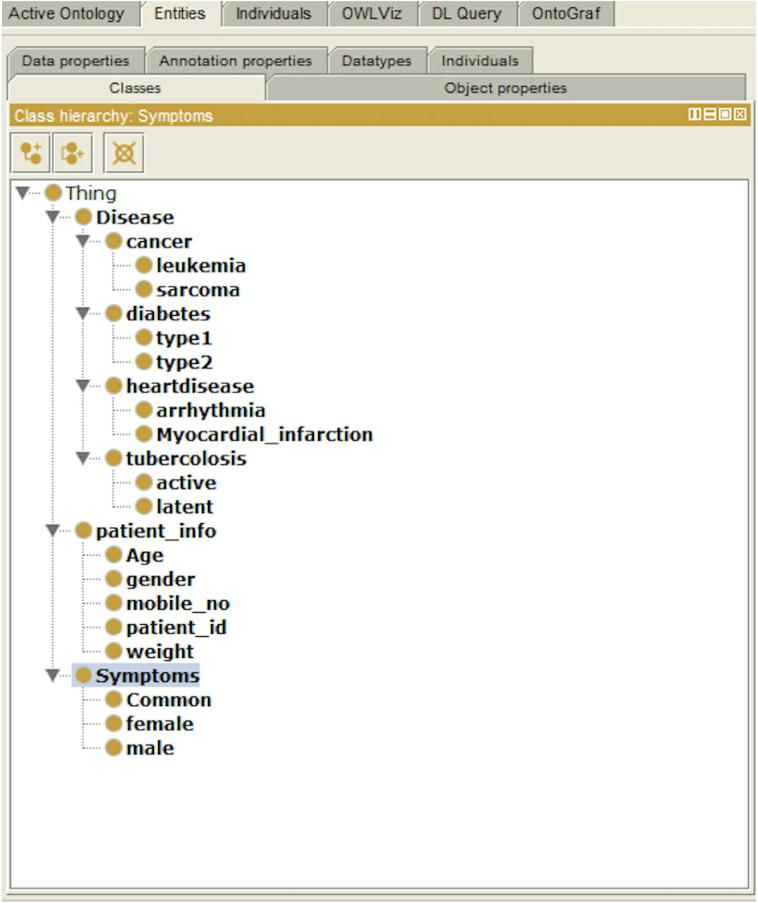


FIG 3: CLASS HIERARCHY

10

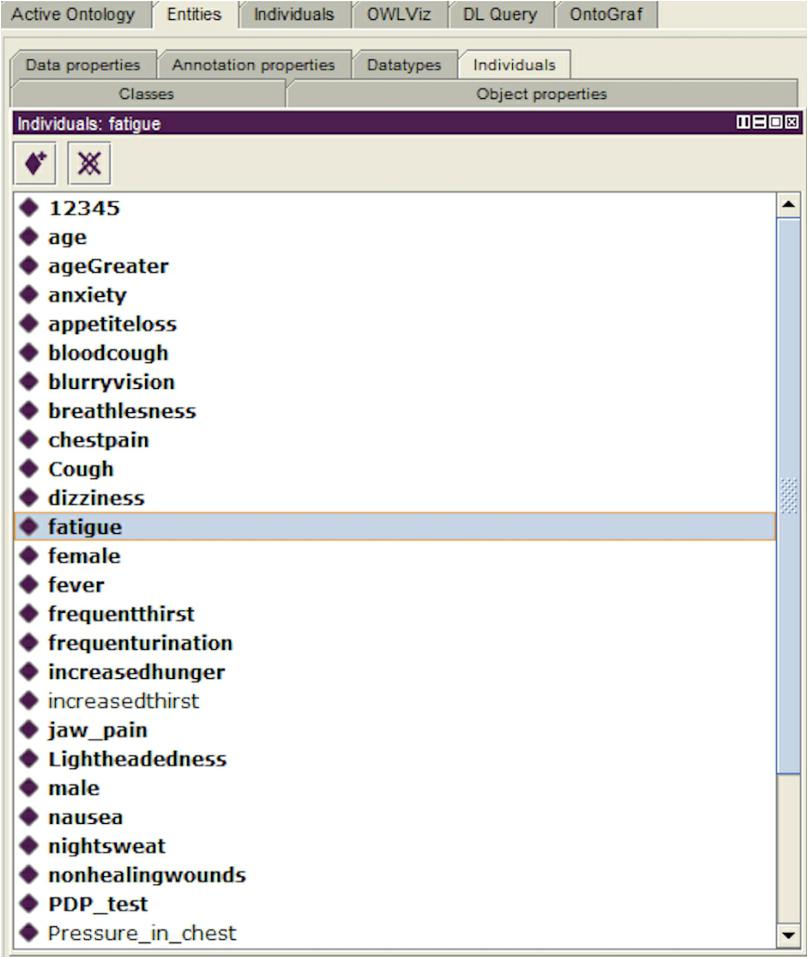


FIG 4: INDIVIDUAL PROPERTIES

Fig. 5: Basic Architecture[4]

12

**Knowledge discovery and rule preparation**

In this phase, the initial data exploration is performed to verify the dataset completeness and missing attributes. Moreover, a technical preparation of the dataset in which pre processing and reformatting mechanisms is performed to meet the requirements of the data mining techniques used within the investigation. Preparing the original dataset is followed by applying a learning system, learning classifier system in this chapter, over the dataset. This allows the creation of a new knowledge base system which it is able to discover compacted, organised, and representative knowledge and to build a prediction model in order to predict future cases, and to deal with generating a readable human-interpretable output to describe the problem effectively. The final step in this phase is to prepare the generated rule set to suit the ontology development engine in the next phase. The low experience of rules indicates that they either match a very small fraction of the dataset, which obviously could be matched by other rules, or by those that were generated late in the training process, which implies that the learning system did not have enough time to decide whether to delete them or approve their fitness. Moreover, the high prediction error of a rule indicates it’s inaccuracy and/or that it has very significant missing information.

13

**Rules**

* IF patient is male And if there is chest pain AND IF there is pain in upper part of the body AND If it last more than 5 mins And if sweating occur And if he feel nausea then he is suffering from myocardial infraction
* IF patient is female And If she having features of ‘RULE 1’ And if there is shortness of breath And IF vomiting occurs Then she is suffering from myocardial infraction
* IF Patient has feature of RULE 1 And IF patient has features of Rule 2’ THEN Recommend “Stress Test“ And Then Recommend ‘Angiogram’ And Then Recommend ‘Echocardiogram’.
* IF patient has slow heart beat AND IF patient feels palpitations AND IF patient experiencing syncope THEN he is suffering from “BARDYCARDIA-Arrhythmias”
* IF patient has fast heartbeat AND IF patient feels palpitations AND IF patient experiencing syncope THEN he is suffering from “TACHYCARDIA-Arrhythmias”.
* IF patient has Arrhythmias THEN recommend Blood Test Then recommend EKG the. Recommend chest-Xray THEN recommend heart catheterisation Then recommend echocardiogram THEN recommend electrophysiologic test.
* IF patient has myocardial infarction THEN consult doctor
* IF patient has bradycardia-Arrhythmias THEN suggest pacemaker
* IF patient has tachycardia\_Arrhythmia THEN perform Vagal manoeuvres
* IF patient doesn’t show positive signs from Rule 9 THEN suggest surgery THEN consult

14

**First Order Predicate Logic**

* U->(PATIENT(M)^ ~PATIENT(F)) v (~PATIENT(M)^ PATIENT(F))
* For All(X,Y) U ^ PERSISTENT\_COUGH^ BLOOD\_COUGH ^ BREATHING \_PAIN ^WEIGHT\_LOSS=> TB
* For All(X,Y) U ^ PDP\_POSITIVE ^ BCG=> FALSE POSITIVE
* For All(X,Y) U ^ PDP\_NEGATIVE ^ AIDS => FALSE NEGATIVE
* For All(X,Y) U ^ PDP\_POSITIVE ^ BLOODTEST\_POSITIVE ^ SPUTUM\_NEGATIVE => Latent TB
* For All(X,Y) U ^ PDP\_NEGATIVE ^ BLOODTEST\_NEGATIVE ^

SPUTUM\_POSITIVE => ACTIVE TB

* For All(X,Y) U ^ LATENT TB => IMAGING TEST
* For All(X,Y) U ^ LATENT TB => (U ^ ~ ACTIVE TB)
* For All(X,Y) U ^ PATIENT(M) ^ HUNGER ^ DECREASED\_SEX\_TIME ^

POOR\_MUSCLE\_STRENGTH => DIABETES

* For All(X,Y) U ^ PATIENT(F) ^ HUNGER ^ FREQUENT\_URINATION ^

UNIARY\_TRACT\_INFECTION => DIABETES

* For All(X,Y) U ^ (PATIENT(M) ^ ~PATIENT(F) ^ DIABETES ^ AGE <=20 => TYPE 1 v (~PATIENT(M) ^ PATIENT(F))
* For All(X,Y) U ^ (PATIENT(M) ^ ~PATIENT(F) ^ DIABETES ^ AGE > 20 => TYPE 2 v (~PATIENT(M) ^ PATIENT(F))
* For All(X,Y) U ^ ((PATIENT(M) ^ ~PATIENT(F) ^ TYPE 1 => (~PATIENT(M) ^ PATIENT(F)

**Explanation**

* This rule states that a patient who is male and his symptoms hunger , decreased sex drive and poor,muscle concludes that the patient is suffering from diabetes.
* This rule states that a a patient who is female and has symptoms hunger ,frequent urination and UTI concludes that the patient is suffering from diabetes.
* The rules state that a patient who is either male or female and has diabetes and is of age less that or equal to 20 is suffering from Type 1.
* The rules state that a patient who is either male or female and has diabetes and is of age is greater than to 20 is suffering from Type 2.
* The rules state that a patient who is either male or female and has type 1 can’t have Type 2 and vice versa.
* The rules state that a patient who is either male or female and has heriditary diabetes of Type 1 has a risk of having Type 1 diabetes.
* The rule states that a patient who is either male o female and has fasting blood sugar less than or equal to 100 is of normal range .
* The rule states that a patient who is either male o female and has fasting blood sugar greater than 100 and less than equal to 125 is of range pre-diabetes.
* The rule states that a patient who is either male o female and has fasting blood sugar greater than 125 is of range diabetes .

15

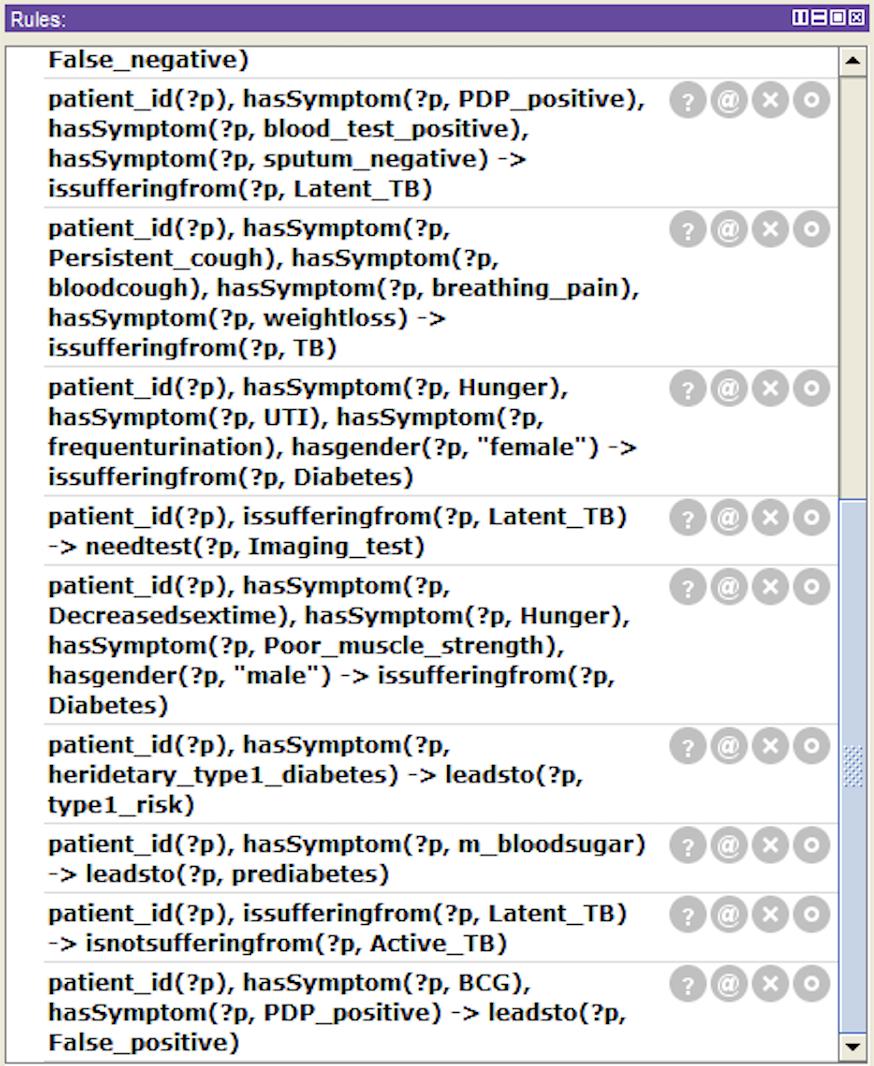


Fig 6: SWRl Rules Created By Converting knowledge based rule

16

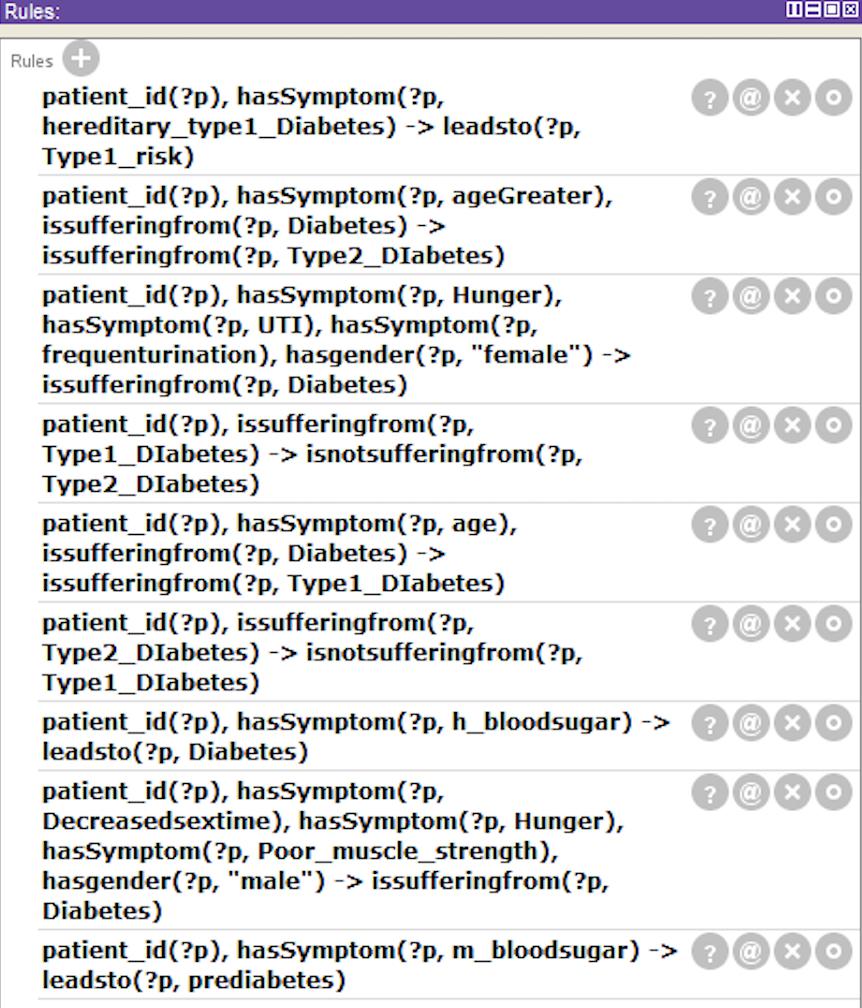


Fig 7: SWRl Rules Created By Converting knowledge based rule

17

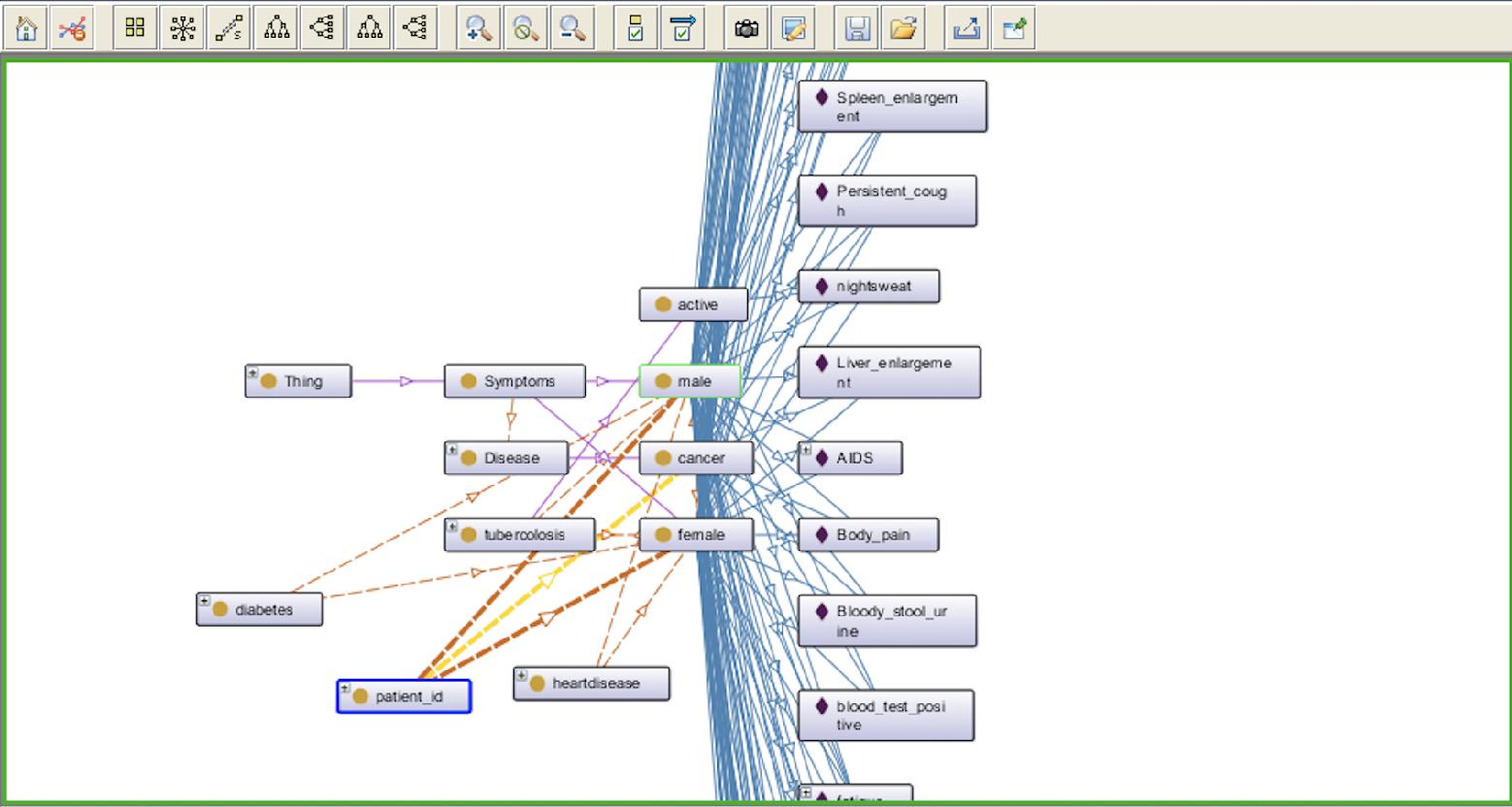


Fig 8: Ontology Graph depicting the relationships between class and individual through properties

18

**CHAPTER 3**

19

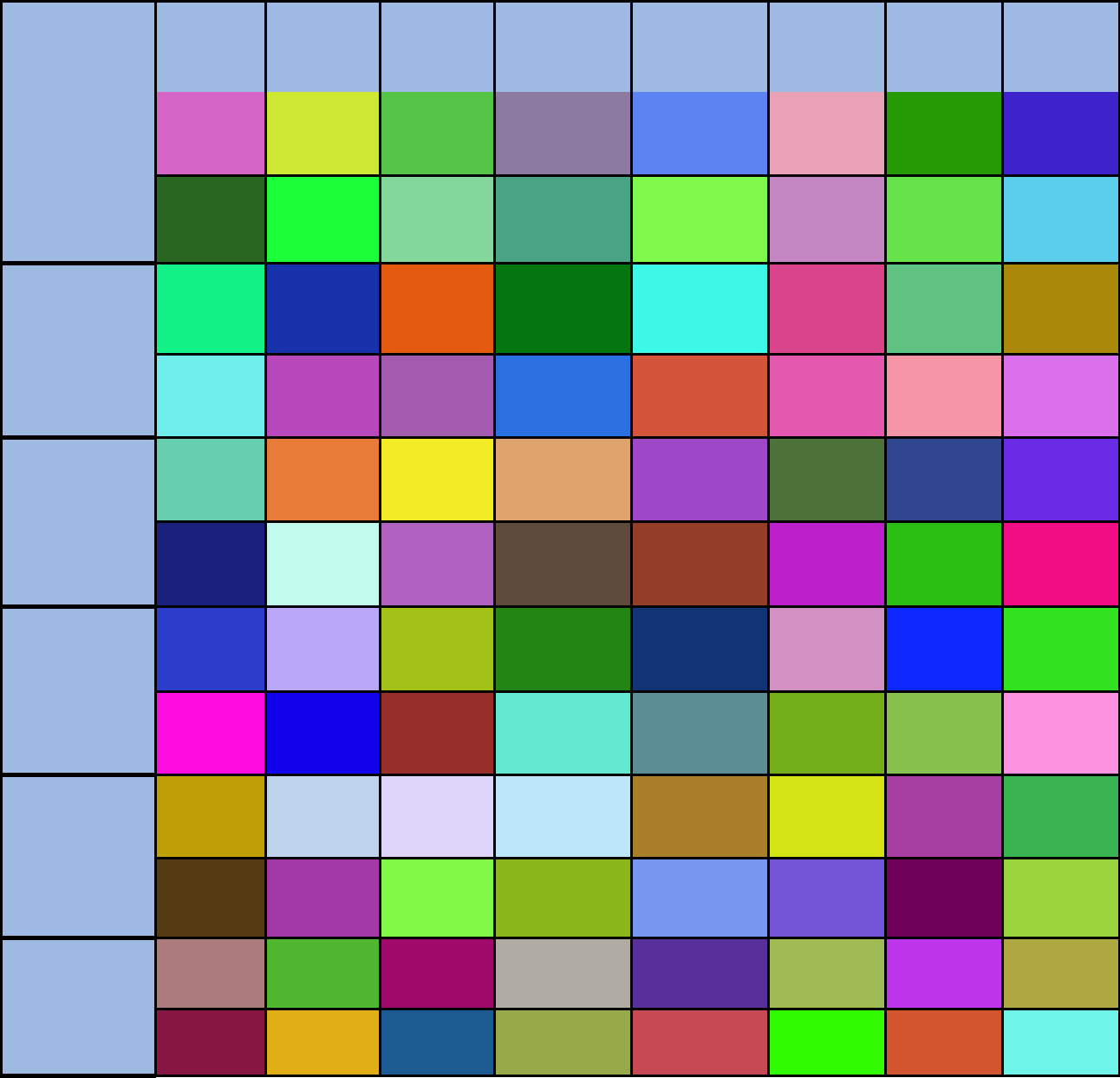
**SOLUTION STRATEGY**

* We determined to collect heterogeneous data from patient in a query manner in the symptoms form.
* This gets further synthesised and mapped with the pre-existing data available from domain experts.
* With the Ontologies, efficiency of pre-processing increases and results can be more precise and accurate.
* Knowledge obtained from the patients will be validated from pre-defined knowledge present in the ontologies.
* Knowledge based rules will be developed using the knowledge extracted from data collected through labs, websites, domain experts and reports.
* Validation of data to be collected will be validated using the reports of the patient concerned.
* Treatment and medication will also be suggested post diagnosis.

20

**GANTT CHART**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ACTIVITY** | 10th Jan-29 | Jan-14th Feb 15th Feb-1s | 2nd March-12t | 13th March-31 | 1st April-16t | 17th April- | 1st May- |  |
|  | Jan | Mar | March | March | April | 30th April | 30th |  |
|  |  | May |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |
| Literature Survey |  |  |  |  |  |  |  |  |



P r o b l e m

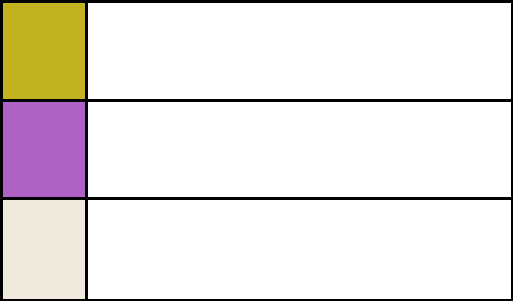
Identification

Analysis

Design

Implementation

Documentation



PROPOSED WORK

ONGOING ACTIVITY

ACHIEVED WORK

21

**REFERENCES**

1. McBride, P.H.a.B., RDF Semantics. W3C, 2004.
2. Mike Dean, G.S., Sean Bechhofer, Frank van Harmelen, Jim Hendler, Ian Horrocks, Deborah L. McGuinness, Peter F. Patel-Schneider and Lynn Andrea Stein, OWL Web Ontology Language Reference, ed. W3C. 2004.
3. Semantics Web and Ontology , Dhana Nadini
4. Knowledge-Based Systems for Development , Priti Srinivas Sajja, Rajendra Akerkar , 2010
5. Building Ontology from Knowledge Base Systems, Faten Kharbat 1 and Haya El-Ghalayini 2

22