Smart Virtual Expert System to Assist Psychiatrists (SVESTAP)

Udara Srimath S. Samarutunge Arachchilage
Faculty of Computing, Department of Software Engineering, Sri Lanka Institute of Information Technology (SLIIT), Malabe, Sri Lanka,
E-mail: samaratunge@gmail.com

Received: 29 July 2017; Accepted: 07 November 2017; Published: 08 January 2018

Abstract—Psychological issues in the world are exponentially growing and the treatment gap is also comparatively high. The main reason would be the shortage of expertise and time-consuming in conventional diagnose process. The main objective of this research is to lower the mental issues treatment gap of professionals or apprentices in the field by creating a virtual expert system to assist psychiatrists. This system diagnoses most common mental disorders such as Depression Disorder, Anxiety Disorder, and Dementia. The proposed expert system can communicate with patients, to identify the current state of the illness. During the conversation, a standard questionnaire is given for the disease verification purpose. The experienced mental health professionals can use this expert system to assist in diagnosing process and the apprentices of the psychology can use this expert system as a training asset.

Index Terms—Psychiatrists, Expert System, Knowledge base, Ontology, Natural Language Understanding (NLU), Natural Language Generation (NLG), Anxiety, Dementia.

I. INTRODUCTION

Health is considered as one of the most vital factors to measure the development of the society, and mental health is significant among them. According to the current statistics, the healthcare industry is one of the top most profitable industries in the world [1]. With the evolution of this scientific knowledge, the healthcare field emerged in areas of medical drugs, treatment methods, and medical equipment and involved with performing research and development activities globally. Even though the knowledge in medicine has been expanding with the time, the patients diagnose process remains as it is since the last couple of years. When diagnosing most mental diseases of patients, the experts would provide a key contribution to determine the exact disease [2]. Based on the gained experience, experts can diagnose mental disorder conditions precisely before proceeding to treatments.

The one major problem in this field is the comparatively low availability of those professional experts and scarce finding them from hospitals in rural areas. Therefore, it would be a challenging task to determine the exact mental disorder for psychiatrists who are newly appointed for the field. Furthermore, erroneously diagnose illnesses of the patient and proceed with treatments would detrimentally impact for the patient and eventually, the whole effort would be in vain. Hence, it is impelled to develop virtual expert system in purpose of assisting for the psychiatrist is key concern. Consequently, construct knowledgebase specific to the psychological diseases can be utilized to pinpoint the exact mental illness.

II. RELATED WORKS

There have been many types of researches and experiments were conducted to develop decision support systems as an attempt to enhance the productivity and efficacy during disease diagnose. Since healthcare domain is a vast area considering immense clinical terms it comprises, and many symptoms and diseases need to be identified, analyzed, and categorized before moving towards treatments. Mainly virtual experts use fed rules and factors to analyze the situation and make decisions but, the accuracy of the decision is the crucial factor to be considered. When analyze the existing expert systems their behavior, and implemented focus is categorically different.

A. MYCIN expert system

MYCIN was an early expert system that used artificial intelligence to identify severe infections, such as bacteremia and meningitis, and to recommend antibiotics, with the dosage adjusted for patient's body weight [3]. Furthermore, MYCIN operated using a simple inference engine, and the knowledge base embedded with rules [4]. MYCIN is one of the well-known programs that embody intelligence and provide data to behave in an intelligent manner. Compared to other artificial intelligent (AI) programs, it was slow in progress and not always in forwarding direction [3].

B. DXplain decision support system

Apart from that, DXplain is a decision support system that developed at the laboratory of computer science at the Massachusetts General Hospital, which has the characteristics of both electronic medical textbook and a
medical reference system [5]. DXplain can provide a list of clinical manifestations such as signs, symptoms, and laboratory examinations for approximately 2000 diseases [5]. Moreover, it provides justifications for each of these diseases and suggests what further clinical information should be collected pertaining to each disease.

C. CADUCEUS medical expert system

In addition to that, CADUCEUS was a medical expert system completed by Harry Popel in University of Pittsburgh in 1980, which took around ten years to build the knowledge base having ability to diagnose up to 1000 different diseases [6]. Both expert systems (CADUCEUS and MYCIN) were developed using the concept of inference engine and incorporate abductive reasoning to deal with additional complexity of internal diseases [6,7].

It is hard to find the expert systems which diagnose most mental illnesses and the above-mentioned expert systems (CADUCEUS, Mycin, DXplain) are not strong enough to recognize patient’s symptoms and determine the disease exactly.

D. Mental Health Diagnostic Expert System (MeHDES)

This is AI rule-based reasoning (fuzzy based reasoning) system developed using techniques such as fuzzy logic and fuzzy genetic algorithm [8], [9], [10]. This fuzzy based reasoning is one approach to construct domain knowledge. It is extended version of semantic web rule and it does not create any inconsistencies for the knowledgebase. Nevertheless, describe the ontology in fuzzy knowledgebase has limited use and fuzzy inference is limited to rules only [11]. As well this MeHDES does not support question and answering approach and system did not act as a separate human counterpart to identify the situation in a conversational manner.

Therefore, the system needs to be more dynamic and it should be further enhanced to response dynamic questions. Hence it is obvious, to develop virtual expert system rule-based reasoning is not only sufficient, whereas understanding user’s question and generate suitable answer is another essential part of it. Therefore, the proposed system fulfills the research goal by performing the role of the virtual human counterpart.

Furthermore, it uses the conversational approach to diagnose the patient and its behavior is highly dynamic when making the decision based on the patient’s response.

III. RESEARCH METHODOLOGY

How this proposed system fulfills the key idea of the research goal is significantly considered during the system design stage. Designing smart virtual expert system is the main research problem and this can be decomposed into four substantial research questions. Refer the following Table 1 represents the defined approach that addresses each research question.

Table 1. Research based question map with defined approach

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Defined approach</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) How to construct the knowledge in expert?</td>
<td>a) Design Knowledgebase</td>
</tr>
<tr>
<td>2) How to diagnose patients?</td>
<td>b) Design question and answering approach to diagnose patients</td>
</tr>
<tr>
<td>3) How to query the knowledge in expert?</td>
<td>c) Design and implement SPARQL queries</td>
</tr>
<tr>
<td>4) How to generate the response?</td>
<td>d) Implement Natural Language Generation (NLG) Technique</td>
</tr>
</tbody>
</table>

According to the high-level diagram depicted in Fig. 1 diagnostic process can be described as follows.

1. The knowledge base comprises necessary knowledge to diagnose the illness, and this is acquired by a domain expert.
2. During this diagnose process, the proposed expert system asks questions from the patient and analyze
the patient’s response using natural language understanding (NLU) technique.

3. This NLU technique analyzes those responses (language: English) and extracts keywords to query the knowledge base.

4. SPARQL queries are used to query the knowledgebase and extracts information based on the given keywords in step 3. Then find most probable disease and send those results to the response formation process.

5. Natural language generation (NLG) technique is used to form the response in human understandable manner.

The Fig. 2 represents interactions of each component in this proposed system.

A. Knowledgebase Construction for SVESTAP

The knowledgebase is the most important part in this SVESTAP to diagnose illnesses. The knowledgebase is constructed by mapping domain experts’ knowledge into an ontology [12]. For this, knowledge acquisition session is conducted with real human experts. Furthermore, to accomplish this knowledge acquisition process professional psychiatrist or panel of psychiatrists need to be interviewed through a questionnaire. This questionnaire should be meticulously planned to fully elicit the knowledge and experience of these experts. Especially, this diagnose process and judgment of diseases based on the symptoms, are performed during the discussion with patients. According to the symptoms of the patient, inference engine infers new knowledge using the embedded knowledge in the knowledge base.

Ontology comprises the logical structure of data in Extensible Markup Language (XML) format and all rules pertaining to that domain can be stored as Semantic Web Rule Language (SWRL) rules [13]. During this construction process classes, object/data properties, instances (individuals) and rules should be designed [14]. Standard books referred by psychiatrists (Eg. Diagnostic and Statistical Manual of Mental Disorder/ Mental State Examination) are used to map those details to Ontology. Fig. 5 illustrates, how classes are linked with properties to construct relationships in the knowledgebase. To implement the ontology Protégé IDE is used and Fig. 3 depicts class hierarchy related to three mental diseases. It can be visualized Onto Graph view as represented in Fig. 4.
B. Understand the Query Using NLU Technique

NLU technique is used to understand the answer given by the user as depicted in Fig. 6. This can be further divided into sub-processes such as (Morphology, Syntax analysis, Semantic analysis, Pragmatics and Discourse analysis) [15, 16]

1. **Morphology**: Initially, the system breaks the long natural language sentence into sub-sentences to derive the meaning of it.
2. **Syntax analysis**: In this process, it realizes the relationships in between sentences through syntaxes. And mainly consider how words are put together to form the correct sentences and what structural role each word has.
3. **Semantics analysis**: In this step, consider how these words are combined to derive the meaning of the sentence. Hence, the system can highlight which keyword is put through the knowledgebase.
4. **Pragmatics**: Sometimes the sentence gives different meaning based on the context. As a result, in this step system considers how sentences are used in different situations and how it affects the interpretation of the sentence.
5. **Discourse**: In here, it considers how current sentence affects to the interpretation of the next sentence. Therefore, the system can keep track of the next sentence to match ordinary keywords which are coming from the user’s response.

NLU sub processes perform the following tasks to elicit keywords from the sentence.

1. Split the sentence into words (Use the delimiter of space to split the sentence into words).
2. Identify the tag of the words (e.g. noun phrases, nouns, verb phrases, verbs, adjective, etc.).
3. Develop a method to identify keywords of the sentence and negation of the sentence.
4. Develop a method to identify the synonyms of keyword related to the answer (when keywords are not matched with the given context).
5. Develop a method to check whether any spellings or grammatical mistakes in the sentence.
6. Send extracted keywords of the sentence to knowledgebase querying process.

C. Query the Knowledgebase and Diagnosing

Keywords are extracted from above section B stage (Understand the user’s question using NLU technique) and it is used to query the constructed domain knowledge to derive the correct disease. Query answering is important in semantic web and several query languages were designed for this purpose such as Resource Description Framework Data Query Language (RDQL), Second generation Resource Description Framework Query Language (SeRQL), and most recently used SPARQL [17]. SPARQL queries can be used across diverse data sources, to verify whether the data is stored natively as RDF or viewed as RDF via middleware [18]. SPARQL contains capabilities for querying required and optional graph patterns along with their conjunctions and disjunctions. The output of SPARQL queries can be results-sets or RDF graphs [18, 19].

E.g. Data in turtle format can be represented as below.

```xml
@prefix svestap: <http://xmlns.com/svestap/0.1/> .
_:a svestap:symptom "confusion and disorientation" .
_:a svestap:symptom "memory loss" .
_:b svestap:disease "Dementia" .
_:b svestap:symptom "palpitation" .
_:c svestap:symptom "sweating" .
_:b svestap:disease "Anxiety" .

Use following query to extract data

PREFIX svestap: <http://xmlns.com/svestap/0.1/>
SELECT ? disease ?symptom
WHERE
{
  ?x svestap:disease ?disease
  ?x svestap:symptom ?symptom
}
```
SPARQL uses “?” to define variables and SELECT clause define returning variables as “?symptom” and “?disease” and WHERE clause define matching criteria. Fig. 7 illustrates sample query that retrieves disease by giving the set of symptoms. In here, there are four result sets fulfilling above criteria and following datasets are given as result set in Table 2.

Table 2. SPARQL Results set of matching criteria

<table>
<thead>
<tr>
<th>Disease</th>
<th>Symptom</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Dementia&quot;</td>
<td>confusion and disorientation</td>
</tr>
<tr>
<td>&quot;Dementia&quot;</td>
<td>memory loss</td>
</tr>
<tr>
<td>&quot;Anxiety&quot;</td>
<td>palpitation</td>
</tr>
<tr>
<td>&quot;Anxiety&quot;</td>
<td>sweating</td>
</tr>
</tbody>
</table>

During the conversation, the proposed expert system asks few questions from the patient and analyze patient’s responses to extract symptoms. From the NLU technique, symptoms were extracted and proceed to the next step of identifying disease category those symptoms belong to. Identify the exact disease based on the given symptoms is another key concern. This goal is accomplished in two stages as follows.

1. First execute SPARQL queries separately for each symptom and fetch all probable diseases for given symptom and categorize them as depicted in Fig. 8.

2. Then find intersection of disease sets to find most probable disease. Refer the Fig. 9 for interaction of diseases.

If there is no disease which has all given symptoms, then proposed system get the disease with maximum number of matching diseases. Then the result will be passed to the NLG process to generate the response familiar to user.
D. Response Formation Using NLG Technique

Response formation is the final stage in the proposed system, after querying the knowledgebase (using SPARQL queries) the new inferred knowledge is given as set of keywords. To comprehend the final results, these keywords must be organized in a human-readable manner [20]. As depicted in Fig. 10, NLG process achieves this by converting keywords into human readable natural language sentences.

Response formation is the final stage in the proposed system, after querying the knowledgebase (using SPARQL queries) the new inferred knowledge is given as set of keywords. To comprehend the final results, these keywords must be organized in a human-readable manner [20]. As depicted in Fig. 10, NLG process achieves this by converting keywords into human readable natural language sentences.

NLG technique is implemented using Java language and separate library called “SimpleNLG” needs to be imported. This comprises inbuilt functions to accomplish the tasks depicted in Fig. 11 (Text planning, microplanning and realization) [22]. Sentence generation can be summarized as per the below three steps:

1. Develop a method to retrieve symptom keywords from the knowledgebase.
2. Develop a method to include required words and form meaningful phrases.
3. Develop a method to generate full sentence based on small sentence-phrases using previously mentioned method.

To generate symptoms in question format for the user, separate xml document has been used to refer them as lexicons. The words in this xml document contains nouns, verbs, and complements (adjectives) refer Fig. 12. Then keys are extracted from xml file using Xpath statements and “SimpleNLG” technique assists you to write a program which generates grammatically correct English sentences as illustrated in Fig. 13.

```java
public String genSentence(String noun, String verb, String adjective, String adverb) {
    String output = "";

    // Initialize the NLG Factory
    NLGFactory nlgFactory = new NLGFactory(lexicon);
    Realiser realiser = new Realiser(lexicon);

    // Create a clause
    PhraseSpec ps = nlgFactory.createClause();
    ps.setSubject(noun);
    ps.setVerb(verb);
    ps.setComplement(adjective);

    // Generate sentence based on tense
    if (tense == 1) {
        ps.setFeature(Feature.TENSE, Tense.PAST);
    } else if (tense == 2) {
        ps.setFeature(Feature.TENSE, Tense.PRES)
    } else if (tense == 3) {
        ps.setFeature(Feature.TENSE, Tense.FUTURE);
    }

    // Use Realiser to generate sentence
    String output = realiser.realiseSentence(ps);
    return output;
}
```

After diagnosing the patient, the proposed system conducts a separate patient evaluation process to determine the intensity of the detected disease. This evaluation is performed by giving a Hypomania Checklist (HCL-32) standard questionnaire [21]. The intensity of the disease can be determined based on the
responses given for the questionnaire. In here, system acts as a separate human-counterpart and provides responses for the user’s questions.

IV. RESULTS AND OUTCOME

This proposed system gives ten questions for the patient to answer and patient can answer them by using natural language sentences as shown in Fig. 14. The given response will be processed by the system and it elicits all symptoms from the response of patient to diagnose the disease. Then based on the given symptoms, it identifies the exact disease that patient is suffering from and formulate it as the response. The Fig. 15 represents, how it generates the response by inferring new knowledge.

1) Question: How many hours do you sleep on a day?
   1. Answer: I sleep around about 6 hours per day
2) Question: How about your memory about past week or so?
   2. Answer: I remember something
3) Question: How do you feel these days?
   3. Answer: I feel good
4) Question: What about your communication skills
   4. Answer: Good
5) Question: How old are you?
   5. Answer: I am 34 years old
6) Question: Do you awake the members of your family at midnight
   6. Answer: Yes I do
7) Question: How about your blood pressure
   7. Answer: It is normal
8) Question: Do you see visual hallucination in sometimes?
   8. Answer: I see rarely
9) Question: Do you hear some unusual voices?
   9. Answer: I here always
10) Question: How about your energy?
   10. Answer: I feel weak

Fig. 14. Sample Q&A process with SVESTAP and Patient

Table 3. User feedback summary

<table>
<thead>
<tr>
<th></th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understanding of the application to use</td>
<td>80%</td>
<td>20%</td>
</tr>
<tr>
<td>Features included</td>
<td>85%</td>
<td>15%</td>
</tr>
<tr>
<td>Usage of the application</td>
<td>90%</td>
<td>10%</td>
</tr>
<tr>
<td>Willingness to use the application</td>
<td>95%</td>
<td>5%</td>
</tr>
</tbody>
</table>

According to the chart depicted in Fig. 16 more than 80% of feedbacks were postive regarding the system usability and willingness to use the application is comparatively high.

V. SUMMARY/CONCLUSION

Nowadays the availability of experts to provide services simultaneously for multiple users would be a challenging task. To fulfill this problem, experimental analysis are conducted in many fields in developing similar virtual expert systems. Development of virtual experts would be an ideal solution but, validity of the made decision is questionable before applying them to real environment. Specially in medical domain, accuracy is the most decisive factor to be considered. As a result, development of expert system to assist real human expert would reduce upcoming risk in the medical domain. As the initial step, I would like to suggest a system that works with a real human expert, which maximizes accuracy factor of the made decision. Therefore, the decisions of disease judgement should always be aligned with actual psychiatrist and the end system response should be frequently evaluated. This system can be utilized for apprentices in the medical field to validate made disease judgement and to identify the root cause before proceeding for treatment.

REFERENCES


Authors’ Profiles

Udara Srimath S. Samarutunge Arachchilage received the M.Sc. degree in Enterprise Application Development from Sheffield Hallam University (SHU) in Sheffield, UK and won the best performer award in 2014. He received his B.Sc. Special honour degree in field of Information Technology (IT) from Sri Lanka Institute of Information Technology (SLIIT) in Sri Lanka in 2010 and B.Sc. degree in IT from Curtin University of Technology, Bentley, Perth, Western Australia in 2009.

He is a Researcher, Lecturer and Software Engineer and currently works as Research Engineer in Department of Software Engineering, Institute of Cybernetics, National Research Tomsk Polytechnic University, Russia and Lecturer in Faculty of Computing, Department of Software Engineering, Sri Lanka Institute of Information Technology (SLIIT) in Sri Lanka. His research interests are Semantic Web, Expert System, Knowledgebase construction, Natural Language Processing, Software Engineering, Human Computer Interaction, Wireless Communication.

He is a member of Computing Society of Sri Lanka (CSSL) and manuscript reviewer in IEEE.