

# Bridging the Vocabulary Gap between Health Seekers and Healthcare Knowledge

T. Ananda Priya<sup>1</sup>, S. Kalaivani<sup>2</sup>PG Scholar, Dept of CSE, Sri Venkateswara College of Engineering and Technology, Thiruvallur, India<sup>1</sup>,Associate Professor, Dept of CSE, Sri Venkateswara College of Engineering and Technology, Thiruvallur, India<sup>2</sup>

**Abstract:** A national survey conducted by the Pew Research Center<sup>1</sup> in Jan 2013, where they reported that one in three American adults have gone online to figure out their medical conditions in the past 12 months from the report time. To better cater to health seekers, a growing number of community-based healthcare services have turned up, including HealthTap<sup>2</sup>, HaoDF<sup>3</sup> and WebMD<sup>4</sup>. They are disseminating personalized health knowledge and connecting patients with doctors worldwide via question answering. These forums are very attractive to both professionals and health seekers. For professionals, they are able to increase their reputations among their colleagues and patients, strengthen their practical knowledge from interactions with other renowned doctors, as well as possibly attract more new patients.

**KEYWORDS:** HealthTap<sup>2</sup>, HaoDF<sup>3</sup> and WebMD<sup>4</sup>.

## I. INTRODUCTION

In many cases, the community generated content, however, may not be directly usable due to the vocabulary gap. Users with diverse backgrounds do not necessarily share the same vocabulary. Take Health-Tap as an example, which is a question answering site for participants to ask and answer health-related questions. The questions are written by patients in narrative language. The same question may be described in substantially different ways by two individual health seekers. On the other side, the answers provided by the well-trained experts may contain acronyms with multiple possible meanings, and no standardized terms. Recently, some sites have encouraged experts to annotate the medical records with medical concepts. However, the tags used often vary wildly and medical concepts may not be medical terminologies. For example, "heart attack" and "myocardial disorder" are employed by different experts to refer to the same medical diagnosis. It was shown that the inconsistency of community generated health data greatly hindered data exchange, management and integrity [2]. Even worse, it was reported that users had encountered big challenges in reusing the archived content due to the

incompatibility between their search terms and those accumulated medical records [3]. Therefore, automatically coding the medical records with standardized terminologies is highly desired. The Main Aim of this Project is to bridge vocabulary gap between health seekers and providers to code the medical records by jointly utilizing local mining and global learning approaches.

## II. SYSTEM ANALYSIS

It is worth mentioning that there already exist several efforts dedicated to research on automatically mapping medical records to terminologies. Most of these efforts, however, focused on hospital generated health data or health provider released sources by utilizing either isolated or loosely coupled rule-based and machine learning approaches. Compared to these kinds of data, the emerging community generated health data is more colloquial, in terms of inconsistency, complexity and ambiguity, which pose challenges for data access and analytics. Further, most of the previous work simply utilizes the external medical dictionary to code the medical records rather than considering the corpus-aware terminologies. Their reliance on the independent external knowledge may bring in inappropriate terminologies. Constructing a corpus aware terminology vocabulary to prune the irrelevant terminologies of specific

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

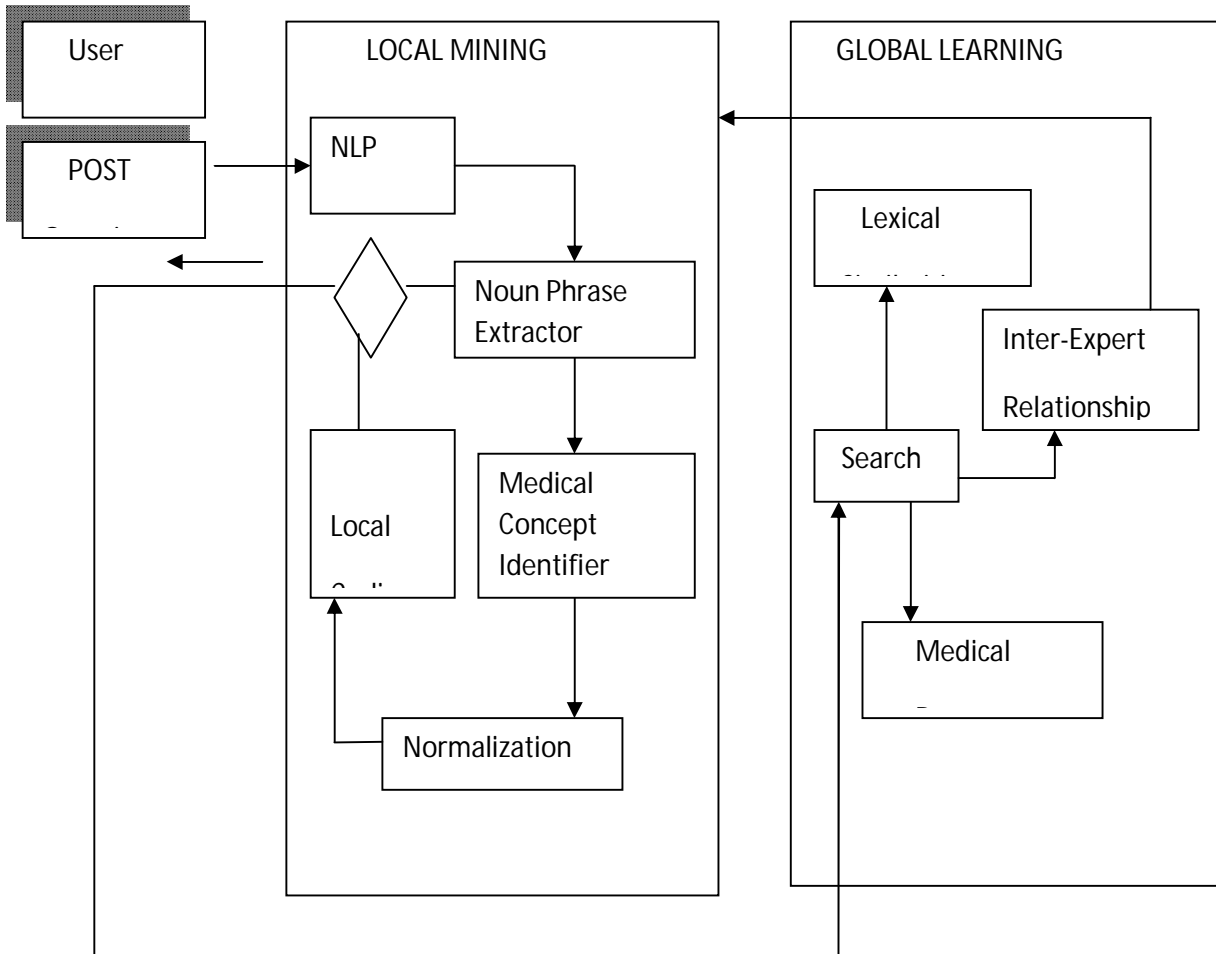
dataset and narrow down the candidates is the tough issue we are facing. In addition, the varieties of heterogeneous cues were often not adequately exploited simultaneously. Therefore, a robust integrated framework to draw the strengths from various resources and models is still expected.

We propose a novel scheme that is able to code the medical records with corpus-aware terminologies. As illustrated in Figure 1, the proposed scheme consists of two mutually reinforced components, namely, local mining and global learning.

The main contributions of this work are threefold:

To the best of our knowledge, this is the first work on automatically coding the community generated health data, which is more complex, inconsistent and ambiguous compared to the hospital generated health data. It proposes the concept entropy impurity approach to comparatively detect and normalize the medical concepts locally, which naturally construct a corpus-aware terminology vocabulary with the help of external knowledge. It builds a novel global learning model to collaboratively enhance the local coding results. This model seamlessly integrates various heterogeneous information cues.

### Architecture Diagram:



# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

## III. SYSTEM DESIGN

### MODULES

- Q and A Application
- NLP Process
- Bridging Gap
- Machine Learning

### MODULE DESCRIPTION

#### Q and A Application

Generally, In Existing Web Applications the Questions posted by the users are answered by the Other User which might result in redundancy and user unbelivability especially for medical related doubts, clarifications and questions. So a medical Experts who can give believable answers should be available all the time which is practically not possible and time Consuming .So we build a Efficient Q and A Scheme which could give Instant Answers Analyzing the Users Objective behind the Question.

#### NLP Process

The User posts Questions for instant answers is processed by a natural language processing technique so that the proper meaning would be revealed. The Nlp Process comprises a several steps . Of which Parts Of Speech Tagging (POST) results in Phrases and Nouns Extraction. The Keywords thus Extracted is subject to Stemming Process which eliminates the Stop words in the sentence and also trims the keyword for Base Word.

#### Bridging Gap

The Proper meanings will be analyzed with a English Dictionary and the Medical Terms will be Normalized based on Domain Specific Knowledge. Medical Terminologies were Collected and grouped so that the checking with the synonyms of keywords could result in Normalization. The Normalized words will be checked for Contradictions with medical terminologies and the related answers will be Queried from Local Mining Database.

#### Machine Learning

Machine Learning in Our Approach is achieved by the use of Local Mining and Global Learning techniques. Local Mining database gets updated by the Global learning datas once user posts a newer Kind of Query to the Answering System. The Global learning Comprises a large collection of Medical Related Resources in its backend which helps to retrieve a related resource to the Query based on terminology keywords. This Search is completely indexed an thus the retrieval time is faster. In case of resource insufficiency the Query and the Question will be left in pending state till a expert arrives.Once Experts reviewed the query the answers not only dispatches to the Medical Seekers and also updates the Local Mining Database for future instant retrieval to the related Query from other Users.

## IV. RESULTS AND CONCLUSION

Coding standards are guidelines to programming that focuses on the physical structure and appearance of the program. They make the code easier to read, understand and maintain. This phase of the system actually implements the blueprint developed during the design phase. The coding specification should be in such a way that any programmer must be able to understand the code and can bring about changes whenever felt necessary. Some of the standard needed to achieve the above-mentioned objectives are as follows:

- Program should be simple, clear and easy to understand.
- Naming conventions
- Value conventions
- Script and comment procedure
- Message box format
- Exception and error handling

Naming conventions of classes, data member, member functions, procedures etc., should be **self-descriptive**. One should even get the meaning and scope of the variable by its name. The conventions are adopted for **easy understanding** of the intended message by the user. So it is customary to follow the conventions. These conventions are as follows:

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

## Class names

Class names are problem domain equivalence and begin with capital letter and have mixed cases.

## Member Function and Data Member name

Member function and data member name begins with a lowercase letter with each subsequent letters of the new words in uppercase and the rest of letters in lowercase.

At the culmination of integration testing, software is completely assembled as a package. Interfacing errors have been uncovered and corrected and a final series of software test-validation testing begins. Validation testing can be defined in many ways, but a simple definition is that validation succeeds when the software functions in manner that is reasonably expected by the customer. Software validation is achieved through a series of black box tests that demonstrate conformity with requirement. After validation test has been conducted, one of two conditions exists.

- The function or performance characteristics confirm to specifications and are accepted.
- A validation from specification is uncovered and a deficiency created.

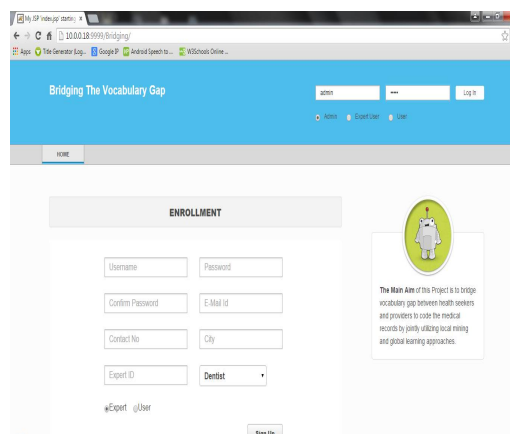
Deviation or errors discovered at this step in this project is corrected prior to completion of the project with the help of the user by negotiating to establish a method for resolving deficiencies. Thus the proposed system under consideration has been tested by using validation testing and found to be working satisfactorily. Though there were deficiencies in the system they were not catastrophic.

## i) USER ACCEPTANCE TESTING

User acceptance of the system is key factor for the success of any system. The system under consideration is tested for user acceptance by constantly keeping in touch with prospective system and user at the time of developing and making changes whenever required. This is done in regarding to the following points.

- Input screen design.
- Output screen design.

## Screenshots:

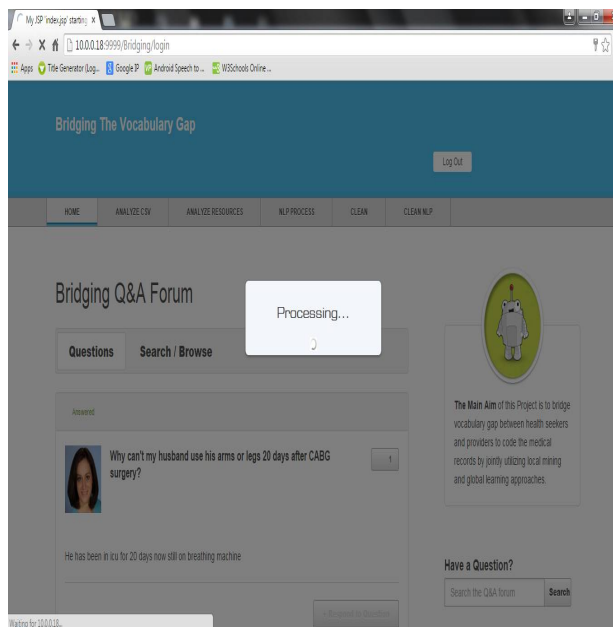
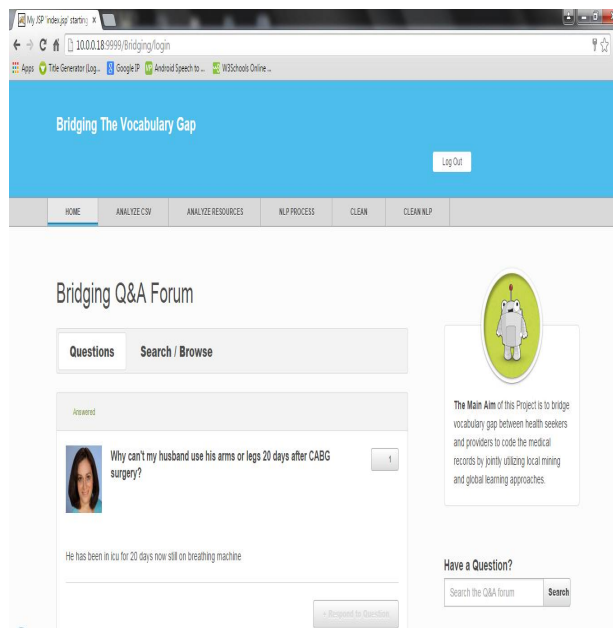


The screenshot displays a web browser window with the URL <http://10.0.0.18:9999/bridging/>. The page title is "Bridging The Vocabulary Gap". The main content area is titled "ENROLLMENT" and contains a form with the following fields: Username, Password, Confirm Password, E-Mail Id, Contact No, City, Expert ID, and Dentist (a dropdown menu). There are also radio buttons for "#Expert" and "@User" and a "Sign Up" button. On the right side, there is a logo of a person and a text box stating: "The Main Aim of this Project is to bridge vocabulary gap between health seekers and providers to code the medical records by jointly utilizing local mining and global learning approaches."

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

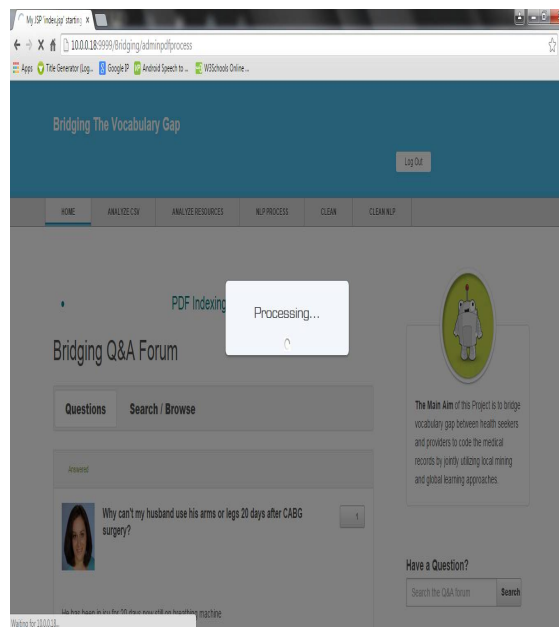
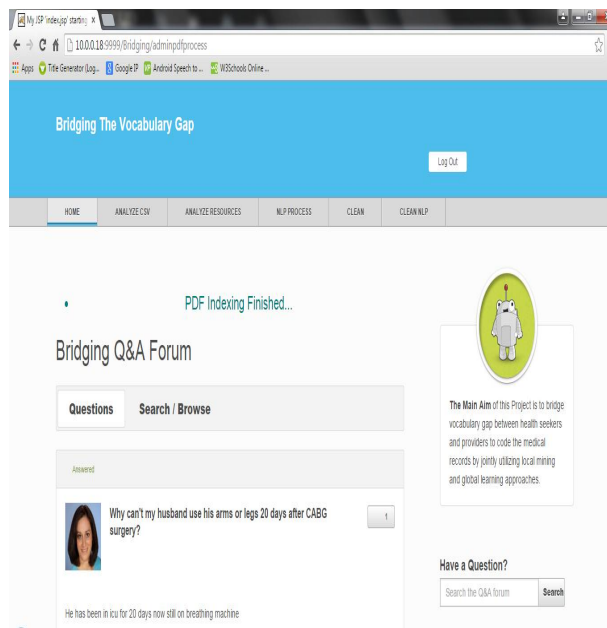
Vol. 5, Issue 6, June 2016



# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

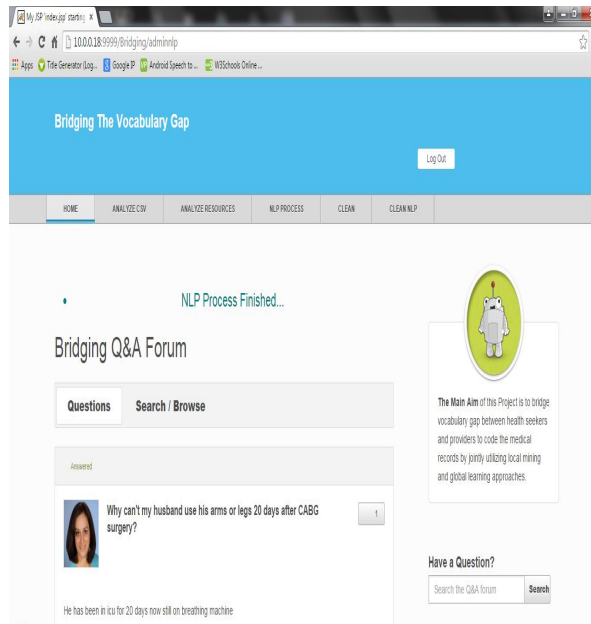
Vol. 5, Issue 6, June 2016



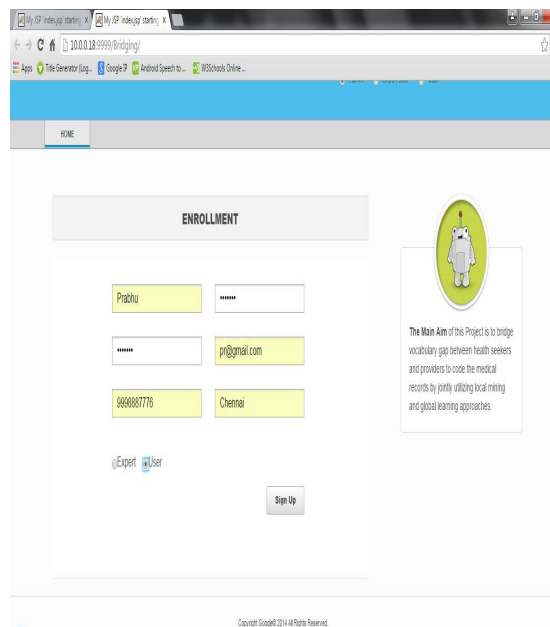
# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016



The screenshot shows a web browser window with the URL 10.0.0.18:9999/bridging/admin/np. The page title is "Bridging The Vocabulary Gap" and includes a "Log Out" button. A navigation menu contains: HOME, ANALYZE Q&A, ANALYZE RESOURCES, NLP PROCESS, CLEAN, and CLEAN NLP. A status message reads "NLP Process Finished...". The main section is titled "Bridging Q&A Forum" and features a "Questions Search / Browse" bar. Below this, a question is displayed: "Why can't my husband use his arms or legs 20 days after CABG surgery?" with a user profile picture. The answer provided is: "He has been in ICU for 20 days now still on breathing machine". A sidebar on the right contains a logo of a robot and the text: "The Main Aim of this Project is to bridge vocabulary gap between health seekers and providers to code the medical records by jointly utilizing local mining and global learning approaches." At the bottom of the sidebar, there is a "Have a Question?" section with a search input and a "Search" button.

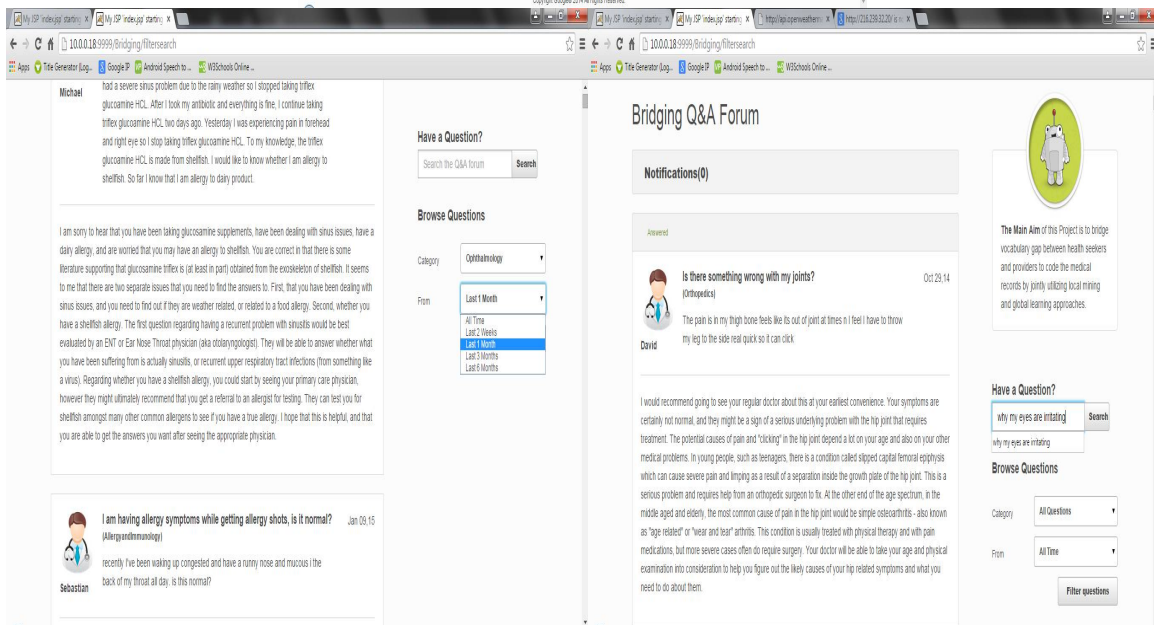
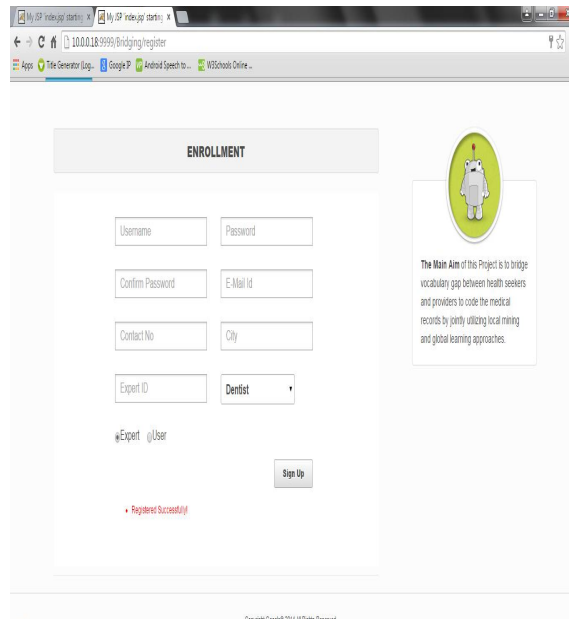


The screenshot shows the "ENROLLMENT" form on the same website. The form includes fields for: Name (Prabhu), Password (masked with asterisks), Email (pr@gmail.com), Phone Number (9998887776), and Location (Chennai). There is also a "Sign Up" button. The sidebar on the right is identical to the previous screenshot, featuring the robot logo and the project's main aim. At the bottom of the page, a copyright notice reads: "Copyright ©2014 All Rights Reserved."

# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016

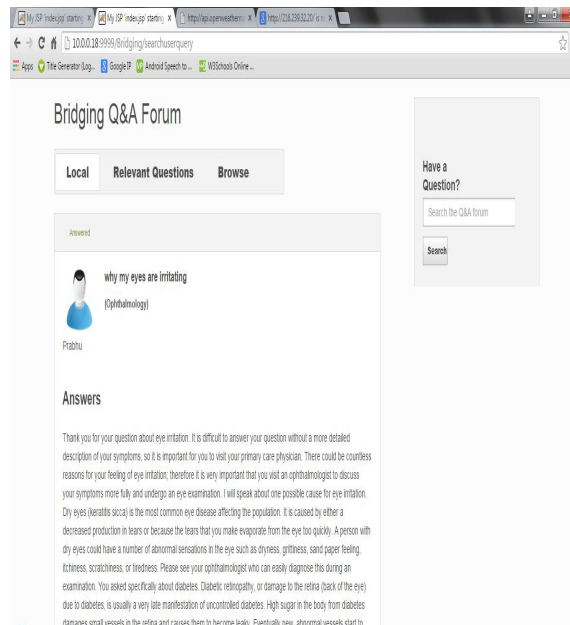
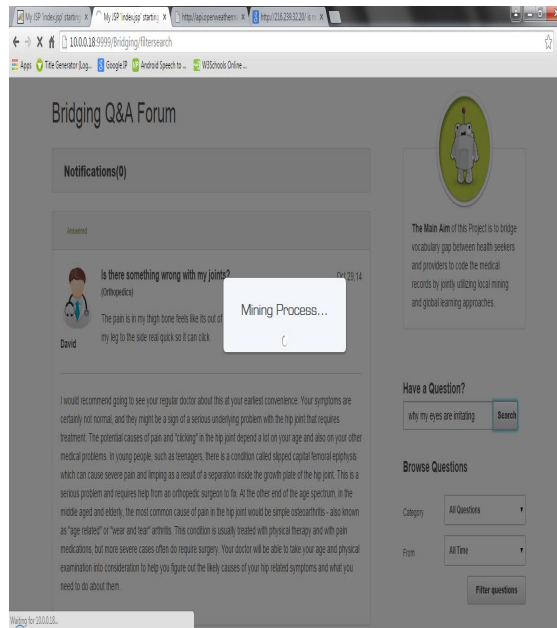




# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

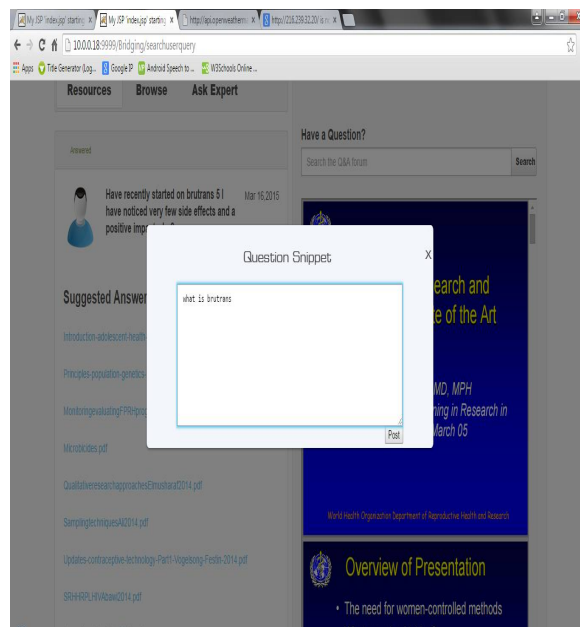
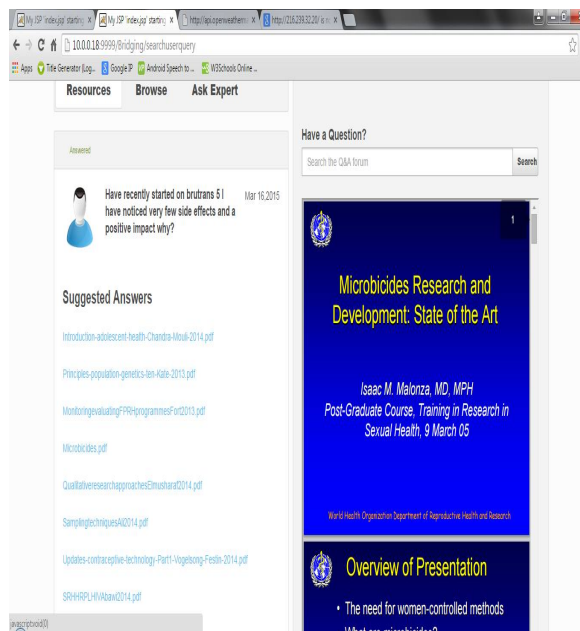
Vol. 5, Issue 6, June 2016



# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

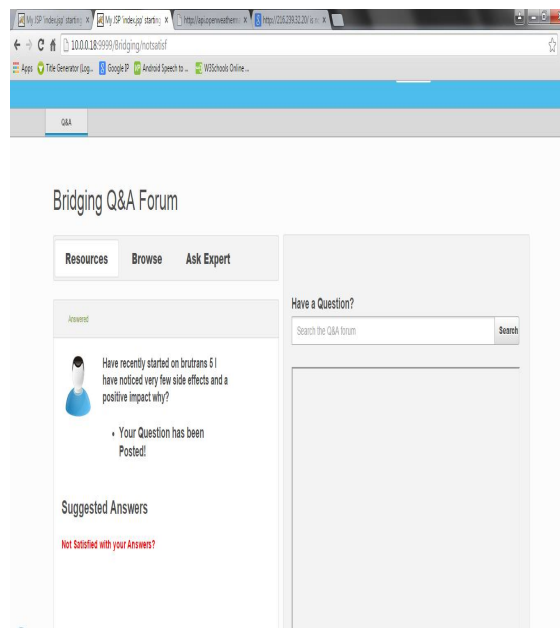
Vol. 5, Issue 6, June 2016



# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

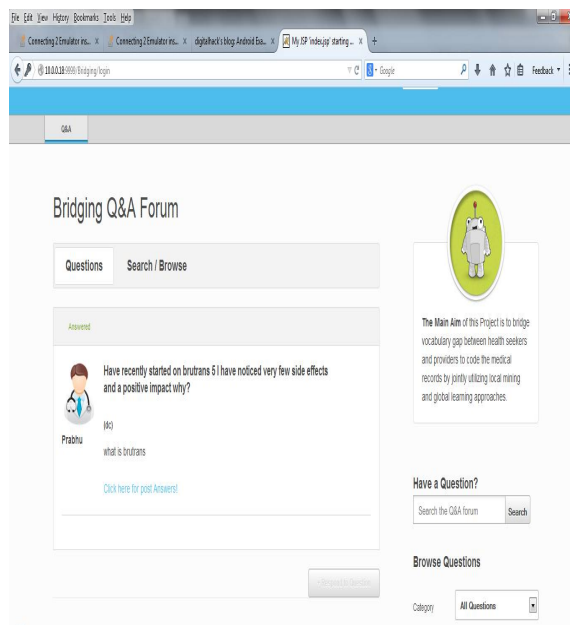
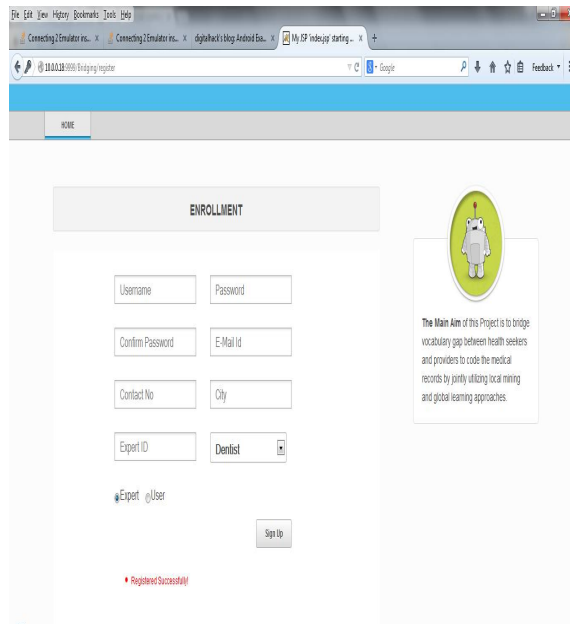
Vol. 5, Issue 6, June 2016



# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

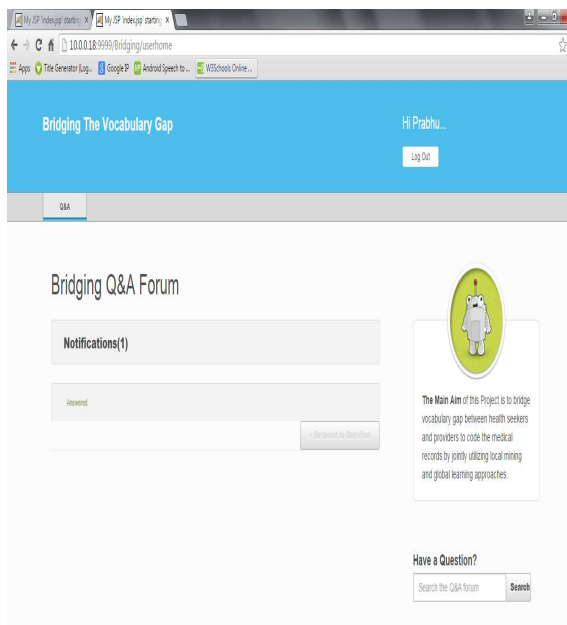
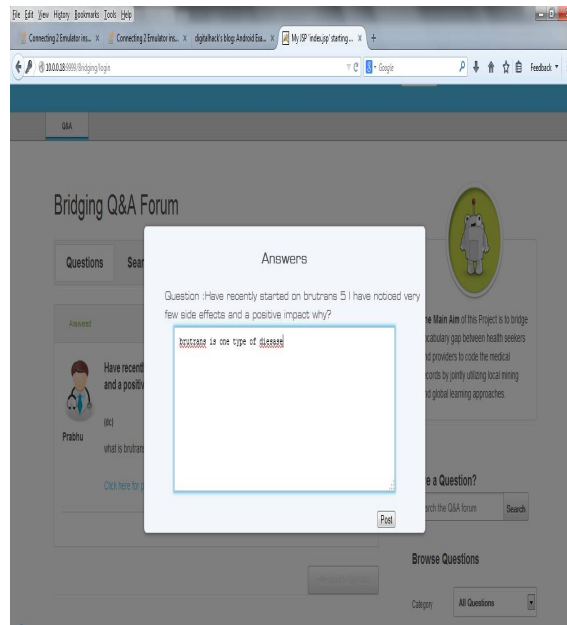
Vol. 5, Issue 6, June 2016



# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

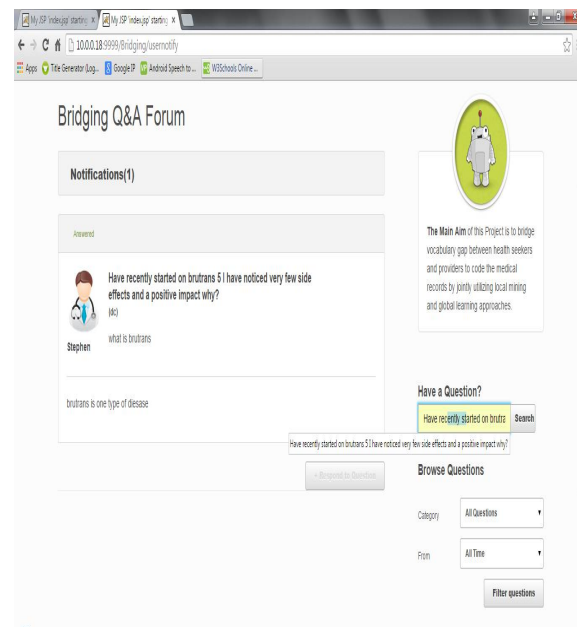
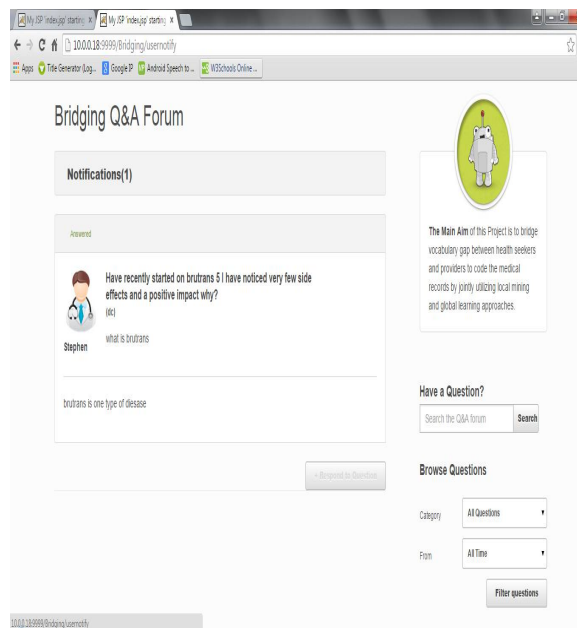
Vol. 5, Issue 6, June 2016



# International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 5, Issue 6, June 2016



## REFERENCES

- [1] A. e-HIM Work Group on Computer-Assisted Coding, "Delving into computer-assisted coding," Journal of AHIMA, 2004.
- [2] G. Leroy and H. Chen, "Meeting medical terminology needs-the ontology-enhanced medical concept mapper," IEEE Transactions on Information Technology in iomedicine, 2001.
- [3] H. Suominen, F. Ginter, S. Pyysalo, A. Airola, T. Pahikkala, S. Salanter, and T. Salakoski, "Machine learning to automate the assignment of diagnosis codes to free-text radiology reports: a method description," in Proceedings of the ICML Workshop on Machine Learning for Health-Care Applications, 2008.
- [4] J. Patrick, Y. Wang, and P. Budd, "An automated system for conversion of clinical notes into snomed clinical terminology," in Proceedings of the fifth Australasian symposium on ACSW frontiers, 2007.
- [5] Q. Zhou, W. W. Chu, C. Morioka, G. H. Leazer, and H. Kangaroo, "Indexfinder: a method of extracting key concepts from clinical texts for indexing," in Proceedings of the AMIA Annual Symposium, 2003.