



ISSN(Online): 2320-9801  
ISSN (Print): 2320-9798

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 5, Issue 1, January 2017

## Artificial Intelligence Powered Bot as a Chatting Agent using AIML

Shruti Deolekar<sup>1</sup>, Chhaya Rathod<sup>1</sup>, Prof. Shubhangi Sonone<sup>2</sup>

B.E Student, Department of Information Technology, D. Y. Patil College of Engineering, University of Pune,  
Pune, India<sup>1</sup>

Head of Department-Department of Information Technology, D. Y. Patil College of Engineering, University of Pune,  
Pune, India<sup>2</sup>

**ABSTRACT:** The mishandle of talk administrations via robotized programs, known as talk bots, represents a genuine danger to Internet clients. Talk bots target famous talk systems to circulate spam and malware. In this paper, we first lead a progression of estimations on a substantial business talk organize. Our estimations catch a sum of 16 distinct sorts of visit bots extending from easy to cutting edge. In addition, we watch that human conduct is more perplexing than bot conduct. In light of the estimation ponder, we propose a classification framework to precisely recognize visit bots from human clients. The proposed characterization framework comprises of two components: 1) an entropy-based classifier; and 2) a Bayesian-based classifier. The two classifiers supplement each other in visit bot detection. The entropy-based classifier is more exact to recognize unknown visit bots, while the Bayesian-based classifier is quicker to identify known visit bots. Our trial assessment demonstrates that the proposed characterization framework is exceptionally compelling in differentiating bots from people

**KEYWORDS:** Artificial Intelligence, Bots, Human Computer Interaction, ChatterBot, Artificial Intelligence Markup Language (AIML).

### I. INTRODUCTION

Web talk is a famous application that empowers genuine time content based correspondence. A huge number of individuals around the world utilize Internet talk to trade messages and examine a wide scope of points on the web. Web visit is likewise a remarkable net- worked application due to its human-to-human collaboration also, low data transmission utilization. Be that as it may, the substantial client base and open nature of Internet visit make it a perfect focus for pernicious misuse. The mishandle of talk administrations via computerized programs, known as talk bots , represents a genuine danger to online clients. Visit bots have been found on various visit frameworks, including expansive business talk systems, for example, AOL, Yahoo!, what's more, MSN , and open visit systems, for example, IRC and Jabber. There are additionally reports of bots in some non-chat frameworks with visit highlights, including web based amusements, for example, Universe of Warcraft Chatbots exploit these online systems to send spam, spread malproduct, and mount phishing assaults.

As such, the endeavors to battle talk bots have concentrated on two diverse methodologies: 1) catchphrase based separating; and 2) human intuitive evidences. The catchphrase based message channels, utilized by outsider talk customers , experience the ill effects of high false negative rates since bot producers much of the time overhaul talk bots to avoid bar lished watchword records. The utilization of human intelligent verifications, such as CAPTCHAs , is additionally inadequate in light of the fact that bot administrators assist visit bots in breezing through the tests to sign into talk rooms . In August 2007, Yahoo! executed CAPTCHA to piece bots from going into visit rooms, yet bots are still ready to enter talk rooms in expansive numbers.

The rest of this paper is organized as follows. Section II covers foundation on visit bots and related work. Section III subtle elements our proposed algorithm and module split up. Section IV depicts various talk bot grouping applications. At long last, Section V concludes the paper and examines bearings for our future work.



ISSN(Online): 2320-9801  
ISSN (Print): 2320-9798

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirce.com](http://www.ijirce.com)

Vol. 5, Issue 1, January 2017

## II. RELATED WORK AND BACKGROUND

### A. CHAT SYSTEM

Web talk is a continuous specialized device that permits online clients to convey by means of content in virtual spaces, called talk rooms or channels. There are various conventions that bolster talk, including IRC, Jabber/XMPP, MSN/WLM (Microsoft), OSCAR (AOL), and YCHT/YMSG (Yahoo!). The clients interface with a visit server by means of talk customers that support a certain talk convention, and they may peruse and join many visit rooms highlighting an assortment of subjects. The talk server transfers visit messages to and from online clients. A talk benefit with a huge client base may utilize numerous visit servers. What's more, there are a few multiprotocol visit customers, for example, Pidgin (some time ago GAIM) and Trillian, that permit a client to join diverse visit frameworks.

Despite the fact that IRC has existed for quite a while, it has not picked up standard ubiquity. This is mostly in light of the fact that its console-like interface and order line-based operation are most certainly not easy to understand. The late talk frameworks enhance client encounter by utilizing realistic based interfaces, and also including alluring components, for example, symbols, emoticons, and sound video communication capacities. Our review is done on the Yahoo! visit arrange, one of the biggest and most mainstream business visit frameworks. Yippee! visit utilizes exclusive conventions, in which the talk messages are transmitted in plain content, while orders, statuses, and other metadata are transmitted as encoded parallel information.

Not at all like those on most IRC systems, clients on the Yahoo! visit organize can't make talk rooms with altered subjects since this component is incapacitated by Yahoo! to counteract manhandle. Likewise, clients on Yahoo! talk are required to pass a CAPTCHA word check test with a specific end goal to join a talk room. This as of late added highlight is to make preparations for a noteworthy source of manhandle—bots.

### B. CHAT BOTS

A chatterbot (also known as a talkbot, chatbot, Bot, chatterbox, Artificial Conversational Entity) is a computer program which conducts a conversation via sound-related or literary techniques. Such projects are regularly intended to convincingly reproduce how a human would carry on as a conversational accomplice, consequently finishing the Turing test. Essentially a chatterbot is a PC program that when you furnish it with a few contributions to Natural Language (English, French ...) reacts with something important in that same dialect.

The term bot, short for robot, alludes to computerized programs, that is, projects that don't require a human administrator. A talk bot is a program that associates with a talk administration to computerize undertakings for a human, e.g., making visit logs. The original visit bots were intended to assist work talk rooms or to entertain talk clients, e.g., test or quote bots. Be that as it may, with the commercialization of the Internet, the principle venture of visit bots is currently sending visit spam. Visit bots convey spam URLs by means of either connections in visit messages or client profile joins. A solitary bot administrator, controlling a couple of hundred talk bots, can appropriate spam connections to a huge number of clients in various visit rooms, making talk bots exceptionally beneficial to the bot administrator who is paid per-click through member programs. Other potential misuse of talk bots incorporate spreading malware, phishing, booting and other malicious exercises. A couple of countermeasures have been utilized to safeguard against the mishandle of talk bots, however none of them are extremely viable. On the server side, CAPTCHA tests are utilized by Yahoo! visit in a push to forestall talk bots joining visit rooms. Be that as it may, this defense gets to be distinctly insufficient as talk bots sidestep CAPTCHA tests with human help.

### C. AIML: ARTIFICIAL INTELLIGENCE MARKUP LANGUAGE

The primary components of the visit bot learning base are portrayed by method for particular AIML labels. The label design encases the client address, while the label format encases the talk bot reply.

AIML (Artificial Intelligence Markup Language) is a XML-consistent dialect that is anything but difficult to learn, and makes it workable for you to start modifying an Alicebot or making one sans preparation inside minutes.

The most imperative units of AIML are:

**<aiml>**: the tag that starts and closures an AIML archive

**<category>**: the tag that denote a "unit of learning" in an Alicebot's information base

**<pattern>**: used to contain a basic example that matches what a client may state or sort to an Alicebot



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirccce.com](http://www.ijirccce.com)

Vol. 5, Issue 1, January 2017

**<template>**: contains the reaction to a client input

There are likewise 20 or so extra more labels frequently found in AIML documents, and it's conceivable to make your own particular supposed "custom predicates". At this moment, a novice's manual for AIML can be found in the AIML Primer.

The free A.L.I.C.E. AIML incorporates an information base of roughly 41,000 classes. Here's a case of one of them:

```
<category>
  <pattern>WHAT ARE YOU</pattern>
  <template>
    <think><set name="topic">Me</set></think>
    I am the latest result in artificial intelligence,
    which can reproduce the capabilities of the human brain
    with greater speed and accuracy.
  </template>
</category>
```

Everything amongst `<category>` and `</category>` is - you got it - a class. A class can have one example and one layout. (It can likewise contain a `<that>` tag, yet we won't get into that here.)

## D. RELATED WORK

Dewes et al. directed an efficient estimation consider of IRC and Web talk movement, uncovering a few measurable appropriate ties of talk movement. 1) Talk sessions tend to keep going for quite a while, also, countless sessions last any longer than Web talk sessions. 2) Visit session inter arrival time takes after an exponential appropriation, while the dispersion of message between entry time is not exponential. 3) as far as message size, all talk sessions are ruled by a substantial number of little bundles. 4) Over a whole session, normally a client gets around 10 times as much information as he sends. Be that as it may, extremely dynamic clients in Web talk and computerized scripts utilized in IRC may send more data than they get. There is significant cover amongst visit and moment messaging (IM) frameworks as far as convention and client base. In general, visit alludes to a framework that backings talk rooms or channels, e.g., IRC, while IM alludes to a framework that backings direct informing and nearness, e.g., Point. Many generally utilized visit frameworks, for example, IRC originate before the ascent of IM frameworks and have incredible effect upon the IM framework and convention outline Liu et al. investigated customer side and server-side strategies for distinguishing and sifting IM spam or spim for short. Be that as it may, their assessment depends on a corpus of short email spam messages, because of the absence of information on spim.

In, Mannan et al considered IM worms, computerized malware that spreads on IM systems utilizing the IM contact list. Utilizing the spreading burn acteristics of IM malware, Xie et al. exhibited an IM malproduct identification and concealment framework in view of the honeypot idea. So also, Trivedi et al. utilized honeypots to dissect system and substance attributes of spim. Despite the fact that not directly identified with talk or texting, Jonathan et al. talk about the issue of socially intelligent malware. Botnets comprise of countless processing resources, which are likewise called "bots." Be that as it may, the use and conduct of bots in botnets are very not quite the same as those of talk bots. The bots in botnets are noxious projects outlined particularly to keep running on traded off hosts on the Web, and they are utilized as stages to dispatch an assortment of illegal and criminal exercises for example, accreditation robbery, phishing, conveyed disavowal of-administration assaults, and different assaults. Despite the fact that having been utilized by botnets as summon and control instruments, IRC furthermore, other visit frameworks don't assume an indispensable part in botnets. Truth be told, because of extensive exertion and advance in identifying furthermore, impeding IRC-based botnets ,the control architectures of later botnets, for example, Zeus, Koobface, and Conficker (or Tempest), are P2P-or HTTP-based instead of IRC-based.

# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirccce.com](http://www.ijirccce.com)

Vol. 5, Issue 1, January 2017

## III. PROPOSED ALGORITHM

### A. DOMAIN OF THE PROPOSED SOLUTION:

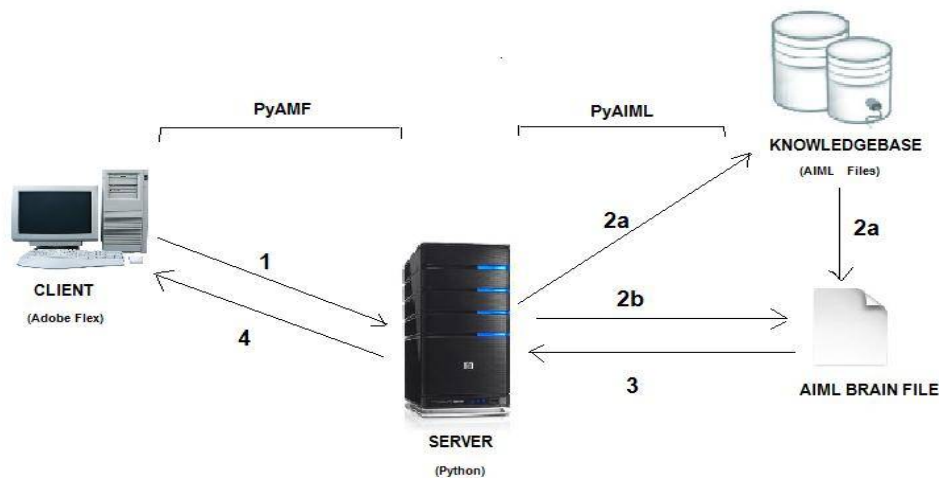
**NLP:** NLP(Natural Processing Language) is a field of computer, AI and PC semantics worried with the connections amongst PC and human dialect

**AIML:** AIML (Artificial Intelligence Markup Language) is a XML-agreeable dialect that is anything but difficult to learn, and makes it feasible for you to start altering an Alicebot or making one without any preparation inside minutes.

### B. DESCRIPTION OF THE PROPOSED ALGORITHM:

- Client makes a demand to the Server.
- Server contacts Knowledgebase/AIML Brain File.
- Server gets the fitting reaction from the AIML Brain document.
- Server advances the reaction to the client.

This algorithmic flow can be pictorially represented as follows:



### C. MODULE SPLIT UP:

The framework incorporates :

#### 1. UI with the accompanying alternatives:

Engage in a discussion, create a modified connection and leave criticism. The last framework incorporates a web interface with the accompanying alternatives: Chat, Login, Leave criticism, include data, overhaul data, erase data, see and erase criticism and see and erase logs. This segments demonstrates the interface of the framework.

#### 2. The Administration interface:

It gives the choices to include inquiries, answers and catchphrases. View, redesign and delete inquiries, answers and catchphrases. View and delete logs. View and delete feedback. The home page of the framework where the client can take part in a discussion is demonstrated as follows:

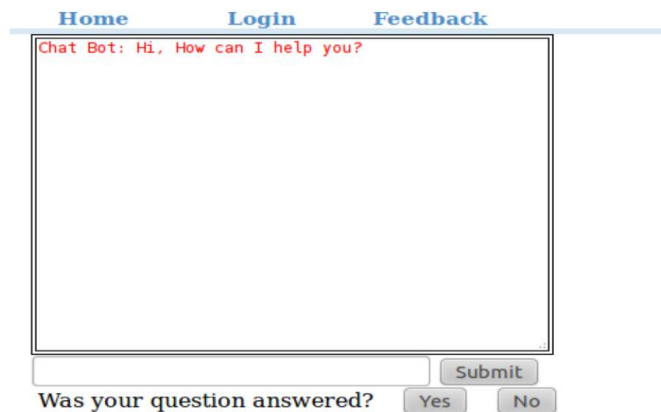


# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirce.com](http://www.ijirce.com)

Vol. 5, Issue 1, January 2017



## IV. SIMULATION RESULTS

We built up a bot that communicates in English this was motivated by our perception that the KorpusGesproke Afrikaans. We mounted models of the chatbots on sites utilizing Pandora bot service<sup>5</sup>, and urged us to assemble the bot. We received three assessment measurements:

- Dialogue efficiency in terms of matching type.
- Dialogue quality metrics based on response type.
- Users' satisfaction assessment based on an open-ended request for feedback.

### 1. Dialogue efficiency metric:

We quantified the effectiveness of 4 test exchanges as far as nuclear match, first word coordinate, most noteworthy match, and no match. We needed to gauge the proficiency of the received learning components to check whether they increment the capacity to discover answers to general client contribution as appeared in table:

Matching Type	D1	D2	D3	D4
Atomic	1	3	6	3
First word	9	15	23	4
Most significant	13	2	19	9
No match	0	1	3	1
Number of turns	23	21	51	17

Table 1. Response type frequency

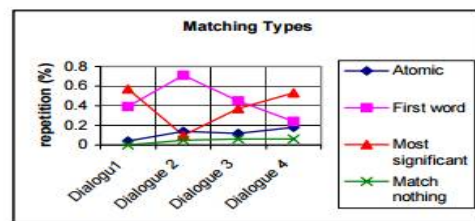


Figure 3: Dialogue Efficiency: Response type Relative Frequencies

### 2. Dialogue quality metric:

With a specific end goal to gauge the nature of every reaction, we needed to group reactions as per an autonomous human assessment of "sensitivity": sensible answer, abnormal yet justifiable, or strange answer. We gave the transcript to an Afrikaans-speaking educator and requesting that her check every reaction as indicated by these classes. The quantity of turns in every discourse and the frequencies of every reaction sort were assessed. Figure 3 demonstrates the frequencies standardized to relative probabilities of each of the three classifications for every example exchange. For this evaluator, it appears that "absurd" reactions are more probable than sensible or justifiable however odd answers.





# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijircce.com](http://www.ijircce.com)

Vol. 5, Issue 1, January 2017

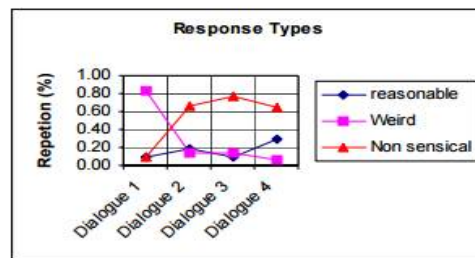


Figure 4: Dialogue Quality: Response type Relative Probabilities

### 3. User's Satisfaction:

The principal models were constructed just with respect to strict example coordinating against corpus expressions: we had not actualized the primary word approach and least frequent word way to deal with include "special case" default classes. Our evaluators found these first models baffling and disappointing: it worked out that few of their endeavours at discussion discovered correct matches in the preparation corpus. Be that as it may, growing the AIML design coordinating utilizing the main word and least frequent-word approaches yielded more great input. Our evaluators found the discussions not so much dreary but rather more fascinating. We measure client fulfilment in view of this sort of casual client criticism.

## V. APPLICATIONS

### 1. A Chatbot as a Tool of Entertainment

The underlying point of building chatbot frameworks was to copy human discussion and delight clients. The main endeavor at building chatbots was ELIZA, which was made in the 60's by Joseph Weizenbaum to imitate a psychotherapist in clinical treatment (Weizenbaum, 1966, 1967).

### 2. A Chatbot as a Tool to Learn and Practice a Language

Clients discovered it an intriguing instrument to rehearse the dialect and appreciated talking, and we reasoned that even with its watchword based coordinating system, a chatbot could be utilized as an apparatus for obscure dialects, where "obscure" means (i) obscure to the chatbot creator/designer, as well as (ii) obscure to computational phonetics, that is, the place there is a deficiency of existing devices to manage the dialects.

### 3. A Chatbot as Information Retrieval Tool

A chatbot could be a valuable apparatus in training, for instance to practice dialect as delineated in segment 5. Knill et al. (2004) found that utilizing a chatbot to answer inquiries will help the instructor to see where understudies have issues, what questions understudies ask, and the produced logs record could be gotten to gauge understudy learning, and understudies shortcomings. The creators built up the Sofia chatbot to help with instructing Mathematics

### 4. Chatbot Assistants in E-Commerce, Business, and different Domains

Shopping colleague is a standout amongst the most influential deals in customary business. Shopping right hand offer assistance in a store, give extra data on items and streamline basic leadership prepare finding a decent that fulfills client's prerequisites and different limitations. (Bogdanovych et al., 2005) Happy Assistant is "a characteristic dialect exchange based route framework that helps clients get to online business locales to discover applicable data about items and administrations" (Chai et al., 2000). The framework is made out of three principle modules: the presentaBand 22 (1) – 2007 43 Abu Shawaar, Atwell tion administrator (PM), the exchange Manager (DM), and the Action Manager (AC).

## VI. CONCLUSION AND FUTURE WORK

This paper first introduces a substantial scale estimation consider on Web talk. We gathered two-month visit logs for 21 diverse visit rooms from one of the top Internet talk specialist co-ops. From the visit logs, we recognized a sum of 16 distinct sorts of visit bots and gathered them into six classes: intermittent bots, irregular bots, responder bots, replay



# International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: [www.ijirccce.com](http://www.ijirccce.com)

Vol. 5, Issue 1, January 2017

bots, replay-responder bots, furthermore, propelled responder bots. Through measurable examination on intermessage deferral and message estimate for both talk bots and people, we found that talk bots carry on uniquely in contrast to human clients. All the more particularly, visit bots display certain regularities in either intermessage deferral or message estimate. Despite the fact that responder bots and replay bots utilize propelled procedures to act more human-like in a few angles, regardless they do not have the general modernity of people.

There are various conceivable zones for future work. In specific, viable sending would bring up a few issues. While our present framework was prepared on information gathered over a large portion of a month, a similar volume of information could be gathered in just a couple of hours framework wide. With an extensive volume of information, the framework could be retrained frequently, and old preparing information would be matured out. Albeit maturing strategies utilized for spam sifting, for example, micro grooming and exponential maturing are pertinent for this review, additionally research is expected to decide the best approach.

## VII. ACKNOWLEDGEMENT

The achievement and nal result of this venture required a considerable measure of direction and assistance from many individuals and weare to a great degree lucky to have this up and down the consummation of my venture work. Whatever we have done is just because of such direction and help and we would not neglect to express gratitude toward them. We regard and express gratitude toward Prof. Shubhangi Sonone for giving us a chance to do the venture work in "Natural Language Processing/Artificial Intelligence" and give all direction and bolster which made us finish the venture on time. We are greatly thankful to her for giving such a pleasant support and direction in spite of the fact that she had a bustling calendar overseeing different affairs. We owe my significant appreciation to our project guide Prof. Shubhangi Sonone , who appreciated our venture work and guided every one of us along , till the finish of our venture work by giving all the essential data to building a decent framework. We would express gratitude toward Prof. Shubhangi Sonone , the Head of our Department for her unlisted consolation and more over for her auspicious support and direction till the fulfilment of our venture work. We are grateful and sufficiently blessed to get steady consolation , bolster and direction from all teaching staff of Department of Information Technology which helped us in fruitful fulfilment of venture work. Additionally, we might want to augment our genuine respects to all the non-showing staff of the office for their auspicious support.

## REFERENCES

1. Dirk Heylen. Talking Head Says "Cheese!" Humor as animpetus for Embodied Conversational Agent Research CHI-2003 WorkShop: Humor Modeling In the Interface
2. Abu Shawer B. and Atwal A. Machine Learning from dialogue corpora to generate chatbots. Expert Update journal, Vol. 6, No 3, pp 25-30 (2003).
3. Steven Birds, Ewan Klein and Edward Lapper. Natural Language Processing with Python by O'Reilly Media, Inc. (2009).
4. (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 6, No. 7, 2015 .
5. R. Mihalcea and C.Strapparava. "Lerarning to laugh(automatically): Computational Models for Humor Recognition". Computer Intelligence, Volume 22, 2006
6. Dumais, S., Banko, M., Brill, E., Lin, J., and Ng, A. 2002. Web question answering: is more always better?.In Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.(SIGIR 2002).Tempere, Finland, pp. 291-298.
7. Kangavari M., Ghandchi S., and Golpour M. A new model for question answering systems. 2008. In proceeding of world academy and science, engineering and technology volume. 2008. Pp. 536-543
8. A.L.I.C.E. Artificial Intelligence Foundation. [Online]. Available: <http://www.alicebot.org/>.
9. A. M. M. Neves, I. Diniz, I., F.A. Barros, "Natural Language Communication via AIML Plus Chatterbots", V Symposium on Human Factors in Computers Systems (IHC 2002), Fortaleza – CE, 387, 2002.
10. A. M. Turing, Computing machinery and intelligence, 1950.
11. LOM (Learning Object Metadata). (2002). Draft Standard for Learning Object Metadata.[Online]. Available: [http://ltsc.ieee.org/wg12/files/LOM\\_1484\\_12\\_1\\_v1\\_Final\\_Draft.pdf](http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf).
12. M. E. Bratman, Intention, Plans, and Practical Reason, CSLI Publications, Program D. Getting Started with Program D. [Online]. Available: <http://www.alicebot.org/resources/programd/readme.html>.
13. R. Asleson, N. T. Schutta, Foundations of AJAX, Apress, 2005. [12] FreeLing. [Online]. Available: <http://garraf.epsevg.upc.es/freeling/>.



ISSN(Online): 2320-9801  
ISSN (Print): 2320-9798

# International Journal of Innovative Research in Computer and Communication Engineering

*(An ISO 3297: 2007 Certified Organization)*

Website: [www.ijircce.com](http://www.ijircce.com)

**Vol. 5, Issue 1, January 2017**

## **BIOGRAPHY**

**Shruti Deolekar** is a student in the Information Technology Department, D. Y. Patil College of Engineering, SavitriBai Phule Pune University. She would be receiving Bachelor of Engineering (BE) degree in 2017 from SPPU, Pune, MH, India. Her research interests are Machine Learning, HCI, Intelligent Systems, Database Management Systems, Algorithms etc. She has sound knowledge of Web Technologies and Computer Programming Languages.

**Chhaya Rathod** is a student in the Information Technology Department, D. Y. Patil College of Engineering, SavitriBaiPhule Pune University. She would be receiving Bachelor of Engineering (BE) degree in 2017 from SPPU, Pune, MH, India. Her research interests are inWeb development, Database management system, testing and analysis of data etc. She is also certified for security fundamentals.