ISSN 2455 – 9091
IJCSET OCTOBER Volume 7 Issue 10
International Journal of Computer Science Engineering and Technology (IJCSET)
https://www.doi.org/10.5281/zenodo.5675496

NATURALL ANGUAGE PROCESS IN GUSING PYTHON

[1] Prathap M, [2] Mr Prabhu S
[1] PG Student, [2] Assistant Professor,
[1, 2] Department of Information Technology,
[1, 2] Hindusthan College of Arts and Science (Autonomous),
[1, 2] Coimbatore, TamilNadu India.

Abstract - Natural Language Processing (NLP) system using Python and Rasp berry Pi. Naturallan guage processing systems have been used in a wide range of tech industries ranging from medical, defense, consumer, corporate. Most NLP systems used currently requires asubsidiaryprocessinghardwareandadefault OS. The system proposedinthispaperisastandalone NLP system which is open source and can be accessed in remote location susing a simple hard ware component. The processes including voice extraction, speech to text conversion, text processing and data base management and speech synthesishavebeenexplainedindetail along with the python modules used tobuildthesystem.Byminimizingthehardwarecomponentsand usingopensource software, a universal, adaptable NLP system has been proposed.

Keywords: [NLP (Natural language processing), Raspberry PI, speech to text conversion, synthesize.]

1. INTRODUCTION

Natural Language Processing (NLP) isanareaofapplicationandresearchthatexploreshowcomputers canbeusedtounderstand and manipulate natural language speechor text to dousefulthings. The foundation of NLP lie in a number of disciplines, namely, computer and information sciences, linguistics, mathematics, electricalandelectronic engineering, artificial intelligence & robotics, and psychology. NLP researche rsaim to gather knowledge on how human beings use and manipulate natural languages to perform desired tasks so that appropriate tools and techniques can be developed. Applications of NLP include a number of fields of study suchasmultilingual and cross-language information retrieval (CLIR), machine transaction, natural language, text processing and summarization, user interfaces, speech recognition, artificial intelligence and expertsystems.

2. LITERATUREREVIEW

NLPresearchersaimtogatherknowledgeonhowhumanbeingste ndtounderstandandusethelanguagesothatappropriatetoolsandt echniquescanbedevelopedtomakecomputersystemsunderstan d and manipulate natural languagesto perform the desired [1] [4] Phonologicalrules are captured through machine learningontrainingsets.Pronunciationdictionariesarealsousedf orbothtext-to-speechandautomatic speech recognition. Sounds as wellaswordscanbepredictedbyusingtheconditional probability theory [7] [6] the inputto a speech recognizer is a series of acousticwaves. The wavesarethensampled, quantified and literally converted to spectralrepresentation.ThemethodofConditionalprobabilityist henusedtoevaluateeachvectorofthespectralrepresentationwith asystemofstoredphoneticrepresentation. Decodingis the process of finding the optimal sequenceofinputobservations. Each successful match is later used in embeddedtraining– amethodfortrainingspeechrecognizers. [2] [3] Python and NLTK Module are mandatory for the following tasks. NLTK moduleis includedasfollows:

Part of Speech tagging and categorizingwords >>> text = nltk.word_tokenize("Andnowforsomethingcompletelydifferent") >>>nltk.pos_tag(text)

Table. 1. Part of Speech tagging andcategorizingwords

Themainintentionofdesigningtheraspberrypiboardistoincreas etheencouragement on learning, experimentationand innovation for students. The raspberry piboard is portable and low cost. Maximum ofthe raspberry pi computers is used in mobile phones [8].

3. CATEGORIZING THECOMPONENTS

Inthissectionwecategorizethenecessaryrequirementfortheproc essashardware and software based upon the properusageofthose parts.

4. HARDWARE COMPONENTS

ThecomponentsneedefforNLPimplementationcanbesummari zedinthefollowingway:

RaspberryPi Unlike CPU, the Graphics ProcessingUnit on the Pi is equivalent to that in a highspecificationmobiledevice.Itcanrun3Dgames and play high-definition video. Withthe right software, a TV and a broadband linkyoucanhavei-Player,YouTubeandothervideos services at your fingertips. Python isintended as an integral part of the ‘standard’teachingtoolkit. An Outlook modelof Raspberry Pi is shownin Figure. 1.

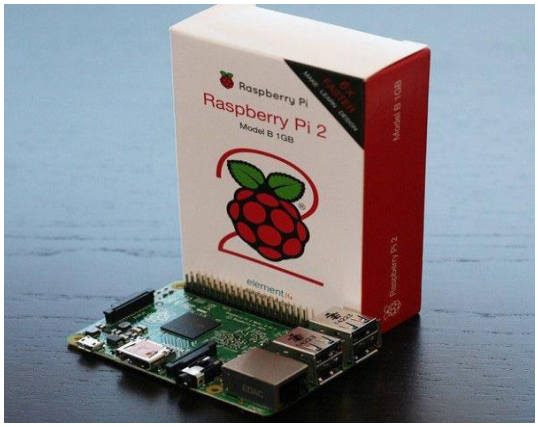

Figure.1.Raspberry Pi2

The Pi comes with 512MB of RAM. Programs are stored on the SD card and the Pi is powere don. They are copied in to the much faster RAM until the computer is turned off and the RAM is cleared. One of the most convenient aspects of Raspberry Pi is that you can convert it fromame diaplayertoa desktop computer just by swapping out the SD card. This is easierthanremovinga laptop's hard disk. A single chip contains the pi's memory, central processing unit, and graphics chip. The version used in the pi is slower than the ones in i-pad and others but it is fast enough to do the job. The architecture of Raspberry Piisshownin Figure. 2.

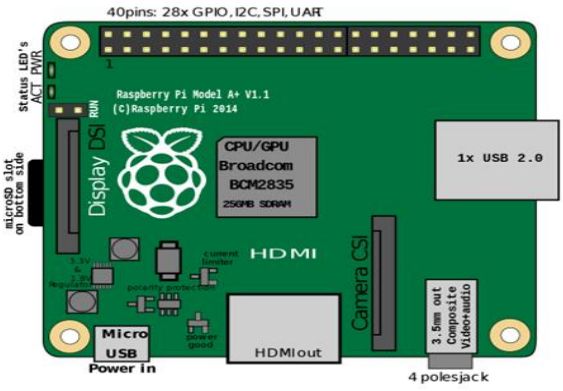

Figure.2.ArchitectureofRaspberryPi

5. MICROPHONE

Ingeneral,a **microphone**isanydevicecapable of recording a voice.It is used as aninput device for inputting the voice.Usuallythe microphone is installed in a CD drive, butin the case of raspberry pi it is downloaded asadriverasitisrequired.LatertheMicrophone is given a source code or a nameforinstanceto becalledduringtheprocess.

SPEECHRECOGNITIONFROMMICROPHONE:

Import speech recognition ask # obtain audio from microphone =sr. Recognizer () withsr. Microphone () as source: printf ("say something!") audio=r.listen (source)

SPEAKER

An operating system is the set of basicprograms and utilities that make our computerrun. At the core of an operating system is thekernel.Thekernelisthemostfundamentalprogram on the computer and lets you startotherprograms. Debian systems use the Linux kernelwhich is a piece of software. FreeBSD is anoperating system including a kernel and othersoftwarein it. However, the work is in progress toprovide Debian for other kernels. The Hurd isa collection of servers to implement differentfeatures that run on top of a microkernel. Likea tower-at the base is the kernel, on top of itareall the basic tools. Nextis thesoftwarethat runs on the computer. At the top of thetoweris Debian. Speakerisusedasanoutputdevicefor sendingouttheconvertedtexttospeechresponse.

6. SOFTWARECOMPONENTS

LINUX

Linux is an open source operating system for computers, mainframes, servers, mobile devices and embedded devices. TheLinux OS includes the Linux kernel as well assupporting tools and libraries. Popular Linux OS distribution sinclude Debian, Ubuntu, Fedora, Red Hat, etc., herewareusing Debain and the reason is specified.

PYTHON

OneoftheadvantagesofPythonisthatitallowsustotypedirectlyin totheinteractiveinterpreter.WecanaccessthePython interpreter using a graphical interfacecalledtheInteractive Development Environment (IDLE). Pythonverycloselyresembles the English language. In this paperthefunctions arecalledusing python.

POCKETSPHINX

Pocketsphinx is a library that dependson another library called SphinxBase. It is alightweightspeechrecognitionengine.Toinstall Pocketsphinx, you need to install bothPocketsphinxandSphinxbase.PocketsphinxcanbeusedinLinux,Windows,MacOS,iPhone and Android. In my paper I am usingthispocketsphinxasaspeechtotextconversion engine. It is converted as an imagefileand extracted for execution.

IBM

The IBM Speech to Text servicesprovidesanAPIthatenablesyoutoaddIBM's speech recognition capabilities to yourapplications.The servicetranscribesspeechfrom various languages and audio formats totext with low latency. This service can also beused instead of pocketsphinx as this providesbothbroadband and narrowband.

RECOGNIZING SPEECHUSINGSPHINX

Try:
Printf ("sphinx thinks yousaid"+r.recognize_sphinx (audio)) exceptsr.UnknownValueError:
Printf ("sphinx could not understand audio") exceptsr. Request Erroeas e:
Printf ("sphinxerror;{0}".format(e))

PROCESSINGTECHNIQUE:

The whole conversion process is classified in to two main sections as follows
Speech to text recognition

Text to speech conversion
Speech to text recognition
Before the process begin swemustinst all the speech recognition module, which

SPEECHTOTEXTRECOGNITION

Before the process begins we must inst all the speech recognitionmodule, which is the Pocketsphinxasofhere.Installationofpocketsphinxiseasyanditr equiresinstallationofthreecomponentsalto gether. Theyarethesphinxbase, pocket sphinx, and pocketsphinx python. prede fined state. The sphinx base is used as abasic layer for the conversion of the speech text. Alanguagemoduleiscreatedinthebeginning which contains all the predefinedsentences.Thetextismatchedwiththemoduleandveri fied.Ifthetextsmatchapositive response is picked from the database. If the input tedtextdoesn't match with the database module the response is searched viaonlinespeechrecognitionmodulesandthematcheddatabasei ssentforfurtherprocessing.Belowisthesystematicrepresentatio noftheinput-output module:

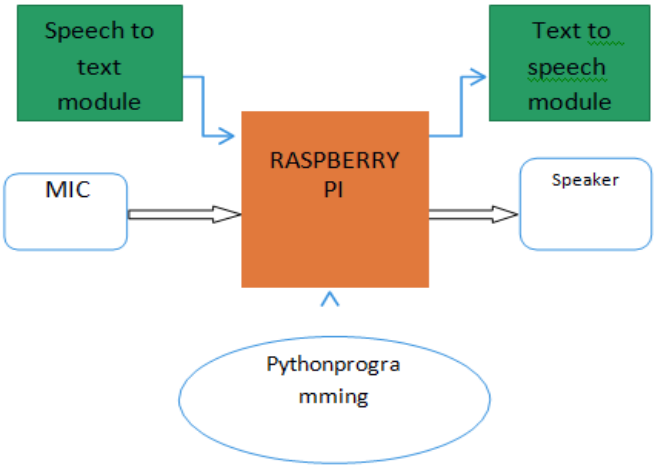

Figure.3. PythonProgrammingBlockDiagram

Texttospeechrecognition
Theconvertedandprocessedtextisnowagain converted to speech. To convert it intospeechamodule calledfestival isused.

Festival is a **free text to speech** tool. Whenwe pass a text file to festival, it converts thecontentsofthe text file into voice.
Installationoffestivalisalsoverysimple.

- sudo apt-get install festivalThisis usedto installfestival.
- TryoutFestivalwith:

CONCLUSION
Natural Language System sarevery complicated to design. NLP's futures will beredefinedasitfacesnewtechnologicalchallengestocreatemor euser friendly systems. It is also forcing NLP more towardsOpenSourceDevelopment.IftheNLPcommunityembr acesOpenSourceDevelopment,itwillmake NLP systems less proprietary and therefore less expensive [9]. Few of the applications of NLP are Personal assistant Assistance for autisticchildren Autism is a life-long developmental disabilitythataffectshowpeopleperceivetheworldand interact with others. Autism is a spectrumcondition. Autistic people hear and feel theworlddifferentlythantheothers.ThisparticularmoduleonNL Pfocusesonprovidingassistanceespeciallyfortheseautistic children. The future enhancement ofthis project will be a major hands-on projecton complete revival from the disabilities, thusproviding them a chance to live a life of theirownchoice. Chatbots-Chatterbotsusenaturallanguage processing to simulate conversationswith users. Web sites are beginning to installchatterbotsasWebguidesandcustomerserviceagents Assistanceindirectingpeopleatpublicrelated places.

REFERENCES
[1]. Natural Language Processing Gobinda G.ChowdhuryDept.ofComputerandInformationSciencesUniv ersityofStrathclyde, Glasgow G1 1XH,UK.
[2]. Ranjit Bose" Natural Language Processing: Currentstate and futuredirections", International Journal of the Computer, the Internet and Management Vol.12#1,January – April, 2004.
[3]. Natural Language Processing Techniques Appliedin Information Retrieval Analysis and Implementation in Python, Tulika Narang, International Journal of Innovations & Advancement in Computer Science IJIACSISSN 2347 – 8616 Volume 5, Issue 4 April2016.
[4]. Bharati A., Chaitanya V andSangal R,"NaturalLanguageProcessing:APaninianPerspective",Pren ticeHall ofIndia,1995.
[5]. Liddy, E. D. In Encyclopedia of LibraryandInformationScience,2ndEd.MarcelDecker,Inc.
[6]. J. All an. Perspectivesonin for mationretrieval and speech. In Information Retrieval Techniques for Speech Applications:LNCS2273, pages1–10, 2002.
[7]. U.S.Tiwary, Tanveer Siddiqui" Natural Language Processingand Information Retrieval" OxfordHigherEducation2008.
[8]. Raspberry Pi Technology Ms.SejalV. Gawande, Dr.Prashant R.Deshmukh, Volume 5, Issue 4, April 2015 ISSN: 2277128 X International Journal of Advanced Research in Computer Science and Software Engineering Research Paper.
[9]. AVinithaetal, Int.J.Computer Technology& Applications, Vol4(5), 751-758,"Natural Language Processing" Preeti and Brahmaleen Kaur Sidhu.
[10]. http://en.wikipedia.org/wiki/Natural_language_processing.
[11]. http://www.doc.ic.ac.uk/~nd/surprise_97/journal/vol1/hks/intro.html.
[12]. http://www.doc.ic.ac.uk/~nd/surprise_97/journal/vol1/hks/techiq.html.