



VOICERAFT: Voice Enabled Active Web Browser with Integrated OCR for Visual Impaired Users

Sandip N. Vende

PG Scholar, Dept. of CSE
JDCT, Indore (M.P), India

Asst. Prof. Krishnakant Kishore

HOD, Dept. of CSE
JDCT, Indore (M.P), India

Abstract: *The many of people from our population are visual debilitated or visually impaired persons and they are not able to get information from massive source of information that is Internet. The way is accessible to such individuals is just a printed books with Braille And it is extremely restricted source of information. This paper describe the development of the Voice enabled smart browser with integration of the Optical Character Recognition System using various technique like Web page Summarization, Speech to Text, Text to Speech and TAF (Text to Audio File) module with advance TAF module that is capable of extracting direct text from PDF Files and transform that text to Audio playable file in (.WAV) format.*

The design of entire system is structured for such individuals that are visually impaired so by caring number of aspects our solution provides a Novel Browser Application that is developed to reduce a maximum difficulties that is generally faced by such people during the Surfing

Index Terms: *Talking Internet Browser, Speech Recognition, Speech Synthesis, Active Web Browser, Smart browsers, Voice enabled Browser.*

I. INTRODUCTION

Surfing web is a fascinating and compelling method for learning, Expressing and Communicating with world .But til our present day age Facilities are not suitable for visual blocked or visual debilitated individual's .The such people are just getting information from Braille coding and some assistive advancements like Braille Embosser print outs or refreshable Braille Display.

The several association are currently a days working for to develop different advancements that will diminish the endeavours of visual disabled individual to get information from PC or web . A number of them are outlined exceptional highlighted programs for such people, yet till visual impeded clients are confronting numerous challenges to handling such complex Systems.

Present frameworks are simply concentrating on just Text Summarization method and building a pleasant visual environment to decrease some inconvenience of low vision alongside outsider Screen Reader Programs that are utilized to peruse out the whole Computer Screen.

The our answer give an extraordinary interface to such individual by pleasant visual environment suitably outline for low vision persons furthermore centred around voice base connection of human and machine. Thus it means that utilization of voice orders empower client to associate with machine straight forwardly without utilization of keystrokes and mouse operation .The PC will likewise react to the client with voice output. The composed arrangement does not require an outsider Screen perusing projects to peruse out the PC screen.

The present programs are not channel the page data independently by which visual weakened client can make sense well over it, on the grounds that web pages are rich with visual impacts, bunches of illustrations and movement by considering this situation our answer gives a compelling parsing of HTML pages and channels the enormous content data freely from interactive media pages.

Our Framework is additionally fit for getting knowledge from diverted content and read loudly in virtual sound space with TTS .Whereas the present highlighted framework are utilizing Screen reader system to read out whole page however when we see through a page it contains a considerable measure pointless data and Screen reader program can't concentrates on particular required content so it read out all superfluous data. By channelize the data the framework decreases time of client to read out the web page

As we know that Visual impaired mean individual with low vision so a number of them have the capacity to peruse out the content straightforwardly from PC screen with some trouble of perusing little size content for that frameworks give spot zooming feature to customer to zoom out the content for this framework distinguish the cursor area and clasp the close-by territory and performs an optical zooming over that region so by this client can have the capacity to peruse out little size anyplace from the page.

The operation of spot zooming performed by simply giving voice order to System and this kind element is not included by present framework.

The framework likewise incorporate the feature of whole page zooming this element will zoom in the entire page by some predetermined rate and it has also the capacity Zoom out the whole page by giving voice summon for that .This

element is useful for the general population who may have an issue of low vision .This element can naturally expand or diminish the extent of content or pictures on page as we need.

The framework likewise coordinates an Optical Charecter Recognition method to perceive content from the printed source .The joining of OCR will help to such individual to peruse out the content from printed material that is not yet conceivable to these people .The framework can take the live pictures from camera and incorporated OCR plugin has the capacity extricate content from pictures taken from live camera .The accuracy of the different OCR is not be generally so precise but rather our incorporated plug-in is able to do simply giving idea of what imprinted on source with normal exactness.

The present system are only prepared for to peruse Text To Speech (TTS) means they are only fit for examining the given content read out loudly. But, Our structure incorporates the TAF module that have the capacity to Transform the Text to Audio Playable File.

PDF records are utilized by a great many individuals over the world. The convey ability is fundamental element of this archive for this element a significant number of Editors ,Authors utilizes a PDF documents .So these E-Book are impressive huge wellspring of data for the normal individuals however with Visual disabled persons PDF are not a suitable formats. For this our proposed arrangement incorporated with Advance Text to Audio File (ATAF) Module . This module will extract accessible text from PDF to Audio File further it will be played or listen by Audio playing gadgets.

II. LITERATURE SURVEY

Speech Recognition System is convert Text to Speech by analysing and processing text using Natural Language Processor and Digital Signal processing [9] .The natural language processing involves text normalization ,Text Linearization and phonetic Analysis, its role is generate phonetic for text .whereas Digital Signal Processing receives transformed symbolic information from NLP and convert it into speech output.

Synthesis system based phonetic connection for the info content .When the framework gets it information in content arrangement , first it change over that content into phonetic interpretation utilizing tenets characterized for changing over expression to sound . The paper [3] propose the TTS framework that chooses the recorded phoneme units (PUs) from database and alters the length of time as indicated by the standard in light of spelling utilizing Time Domain Pitch Synchronous Overlap-Add (TD-PSOLA).

Microsoft has doing part of exploration for enhancing Speech Recognition Systems that are utilized alongside the different adaptations of Windows OS like Windows XP and now in Windows 10 likewise. Microsoft is composed an incredible cluster of Speech Application Programming Interface (SAPI) for the Windows Application .There are a several versions of SAPI is released by Microsoft company .The SAPI was presented in 1995 [10].

The API provided by the Microsoft allows developers to use speech synthesis engine in windows application environment. The API wrapped with the code of dot Net framework is provide the proper interface of Application with TTS engine the paper [5] describes the SAPI overview and shows the Grammar model for increasing accuracy in dot Net Application platform. The Grammar model consists of different element for various purposes. Grammar model is nothing but the a set of rules that are defined for accuracy in prediction of accurate Phrase that will be pronouns by the user. The grammar is specified by specification of W3C in XML format.

The Microsoft Office Document Imaging is tool provided with Microsoft Office to extract the text from Images and scan document. This tool using the Optical Character Recognition with maximum accuracy [8] .The installation of this tool adds (.com) component to dot Net framework and this component is used as plug-in to windows application .This component further processed by the some piece of code written with the code of dot Net platform.

III. PROBLEM STATEMENT

A building such system that will serves to visual hindered individual is truly a major test. There are such a large number of perspective to consider for this individual like how they will begin a PC and how they will interact with computer .By considering their trouble of working with machines, there is requirement for framework that is extraordinarily outline to work with such people. Present framework is not being fit for decreasing their endeavours for examine the web.

When we see the present pattern of site pages they are brimming with visual effects, flashes and videos, So it is very difficult for any program to channel the significant content from these pages. Existing framework just spotlights on content outline and only Text to Speech modules. By considering different blemishes of existing framework there is need of such framework that listen the Voice orders from client rather than mouse and console association. The significant issue is separation of content from mark up of html site pages and speak to the content in virtual sound space with help of TTS.

Issue I:Creating an appropriate interface to Visual hindered client by which they confront slightest challenges by fine GUI and VUI (Voice User Interface) to communicate with Computer.

Issue II: Analyzing the pages to snatch the content in configuration of connections, Headings, Paragraphs exclusively.

Issue III:Processing the content present on Web pages to TTS noisily in important path by which client can bode well over it .

Issue IV: Extraction of content from site pages to change over it in Audio Files with legitimate parameters like Speed, Speaker, Volume and it will be spared in Profile Folder.

Issue V: Implementation of legitimate joining of Optical Character Recognition method to the framework to effectively perceived content from pictures taken out from live Camera.

So in perspective of this above shortcoming this paper proposes System Architecture with assorted modules to gather a such novel web program programming that will help to Visual Impaired persons to open the wonderful gateway of information like conventional persons.

IV. PROPOSED APPROACH

This paper give the usage points of interest of different module that are coordinated in the proposed framework .The thought of entire arrangement is depends on Microsoft Speech SDK .Microsoft speech SDK gave a characteristic and instinctive interface for the engineers and users.

There are two advances are included in TTS engine first is speech recognition and another Speech synthesis. The Text-to-speech (TTS) is normally known as speech synthesis the figure1 demonstrates the structural engineering of Microsoft TTS engine that is utilized as a part of our answer .It figure depicts that how the TTS will get the content information and delivers it into sound. In Text examination stage data content changed over into the arrangement of talked word and it hunt down numbers, times, dates, and other typical representations furthermore changes over condensing to words and layout the how to claim acronyms.

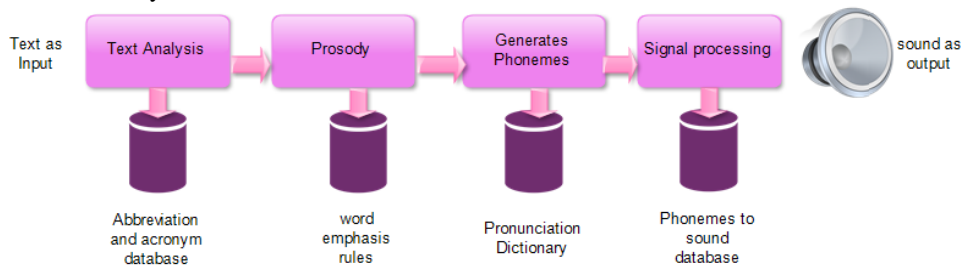


Figure 1: Microsoft Text to Speech engine.

In Figure 2 At the point when client gives speech input to the PC, the sound waves are changed over into computerized sound by the PC's sound card .The basically sound is inspected at 11KHz and 16 bits. The raw sound is initially changed over by the recurrence investigation module to a more valuable arrangement. This includes a great deal of advanced sign handling that is excessively convoluted, making it impossible to portray here. The fundamental test is to extricate the important sound data from the raw sound information.

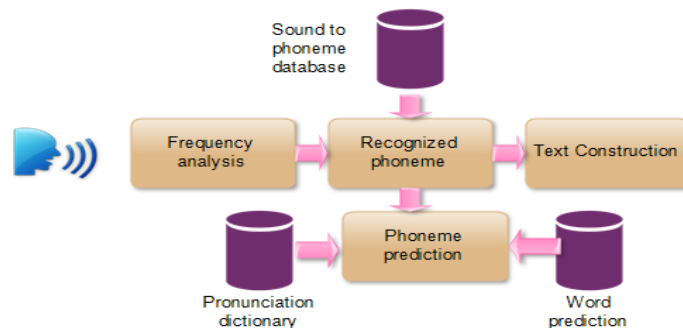


Figure 2 : Microsoft Speech engine.

A. Text Extraction and Summarization

The procedure of content extraction is extremely monotonous assignment ,on the grounds that Web programs organize the different web asset components for the composed page, for example, templates, scripts and images,to present the website page. At whatever point we need remove the data that is required we must need to look at the whole page before extraction of content. Inside the site page content are encompassed inside of the enclosed tag for instance beginning and consummation labels like for straightforward heading "<h1> This is level one heading tag</h1>" so it is extremely hard to independent required content from site page source.

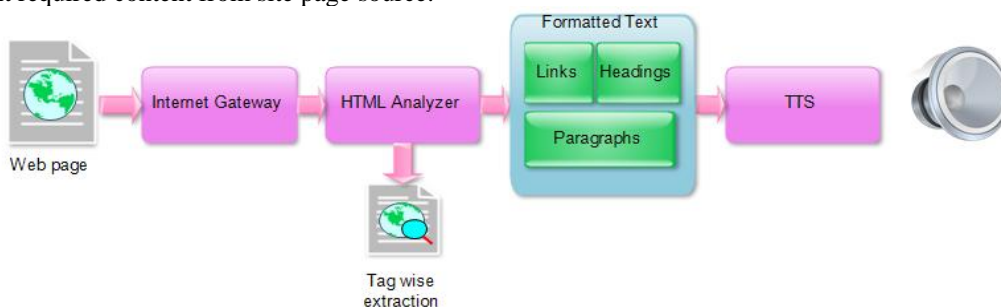


Figure 2: Text extraction module.

There one inquiry is raised with this module that why content extraction is required when it is exhibited in well way with utilization of appealing Style Sheet (CSS) .But when we watch the real section on site page it is encompassed by different graphical components like pictures, promotion, popup so it makes the loads of trouble for low vision pursuer to scrutinize the website page content .By mix of this module we decreased such challenges.

The Figure 2 indicates essential building piece of Text Extraction module. When client requests a page from server the page is given to the Html analyzer after that web page content is arranged with according to label determined .Suppose client gives a voice order for the to concentrate all paragraphs from site page it look at all site page components by label name "p" and makes an accumulation of all sections elements. And give all passage to content management unit to peruse out all passages noisily with speaker. The procedure wise algorithm based dot net framework given underneath.

- a) Initialize the HtmlElementCollection, HtmlElement and HtmlDocument for the analyzing HTML web page.
- b) Set current browser webpage document as HtmlDocument doc for process current web page. Where 'doc' is object of HtmlDocument class of .net framework.
- c) check (WEBBROWSER.Document != null)if it is not null then go ahead or otherwise stop.
- d) Get all elements by tag name using method doc.GetElementsByTagName("p");
- e) Get count of elements that are present in the current web page document.
- f) If count is greater that 0 or count !=null then run loop from 0 to count of available elements.
- g) Get the inner text of current position element by InnerText method.
- h) Return the extracted text to content manager unit to arrange and represent the current extracted text.

B. Voice Command Processing

Proposed arrangement gives a great interface rather than customary interface with mouse and console. The arrangement serves a voice menu to visual hindered clients to pick one a charge from a few orders .After listening voice summon menu client can purport the charge for required operation. Furthermore, for offer affirmation to client the Computer can resound the perceived order to the client so by this client can comprehend that summon is acknowledged by the machine and machine will perform the right activity for it. The voice charge database can't be adjusted by client in light of the fact that the voice order are tied with the peace of code composed at the season of advancement .But arrangement can give the rights to client to include ,erase and redesign voice tag of bookmarked URL database any point and at whatever time. The Figure 3 shows the working of Voice command processing module.

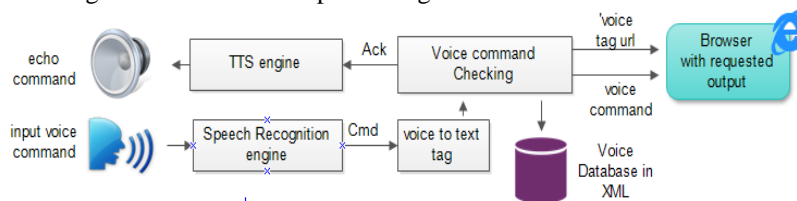


Figure 3: Voice command processing.

At the point when client gives a voice information to the framework then Speech acknowledgment engine make an interpretation of voice to content charge state and confirms it by utilizing voice summon grammar database as a part of xml configuration .It checks the each time information expression is bookmark URL voice label or it is voice order for browser. Suppose it is URL then it ask for a website page in program or on the off chance that it is voice order then it perform its action and shows result on web browser program further it is handled by TTS.

C. Integration of OCR Technique

Optical character recognition technique in propose arrangement utilizing a plugin developed by the Microsoft organization recognized with name Microsoft Office Document Imagination .It first time presented in Office Xp in 2003. The MODI Viewer control is available for any application development tool that support ActiveX controls by including Microsoft Office Document Imaging Viewer Control to the application project. The propose methodology includes a live picture catch from the Web camera by utilizing DirectShow dot net API gave by the Microsoft.DirectShow sdk is multimedia framework and API to perform different operation on media records or Streams. The figure 4 demonstrates interface of DirectShow with MODI where picture caught by Web camera is given to the MODI part to recognized content from the information image. After extraction prepare the content yield is taken care of by TTS engine. The procedure is started by client when he or she puts printed content before camera and gives a voice charge to take a snapshot from live camera and later content is removed from catch picture is done by algorithm itself and further extricated content is perused out by TTS engine.

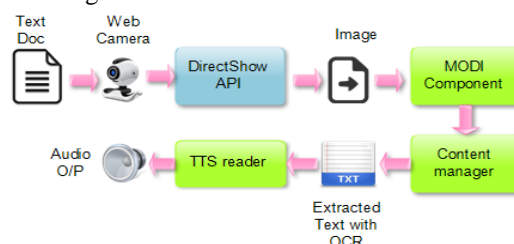


Figure 4: Working of Integrated OCR

D. Text to Audio File (TAF)

TAF is extra utility furnished with the framework that is intended for both kind of client like typical individual and visual weakened individual .In this module client will be proficient for to change over chose content to Sound document in (.wav) format with determined Volume, Reader ,Speed and pitch. The structure likewise fuse Advance TAF module that is able to concentrate content from PDF documents and change over the removed content to sound record. We all know that a many of Authors, Editor and publication house are giving first inclination to distribute their content in PDF records on account of highlight of compactness of this files. So these document are consider as extraordinary wellspring of learning .Thereafter it is not suitable for visual debilitated persons so by utilizing this module they will capable to make their own sound library by changing over PDF to sound records. The Figure 5 shows steps include in content transformation to sound document in TAF model .The chose content or the extricated content from pdf record with use of iText open source API is given to content manager unit and after that by utilizing TTS engine the sound grabber algorithm can makes the sound WAV record.

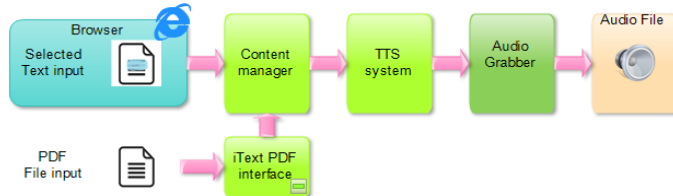


Figure 5: TAF Module using iText PDF interface.

V. RESULT ANALYSIS

The Figure (a) showing screenshot with extracted link after giving voice command ‘facebook’ to the browser ,then system will automatically open the required URL from already bookmarked voice tag XML database.



Figure (a) Main screen of browser after voice command ‘facebook’.

The Figure (b) demonstrates the Screenshot of live character acknowledgment from web camera to concentrate content from ink printed source.

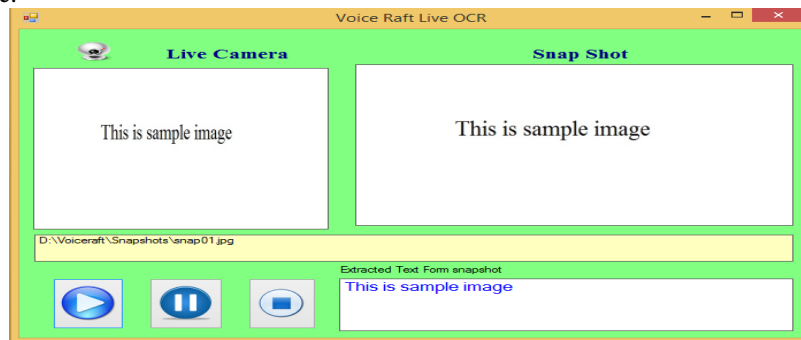


Figure (b) Screenshot of Live OCR module.

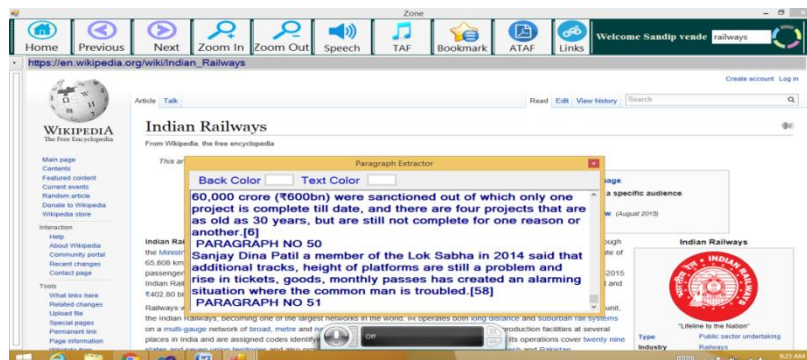


Figure (c) Screenshot of paragraph extraction module with present number of paragraph in web page

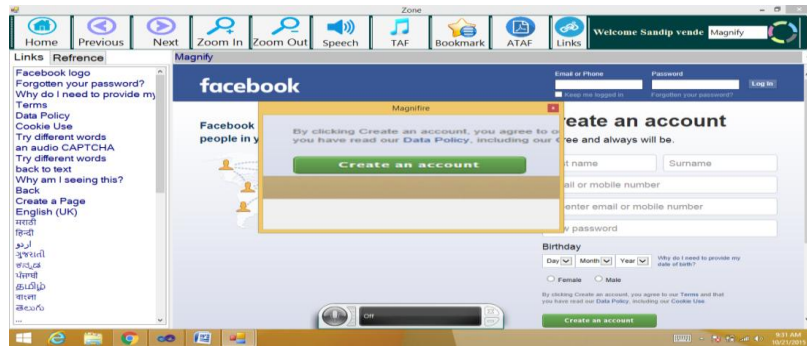


Figure (d) Screen shot of spot magnification utility provided by solution on web page

VI. CONCLUSION

Existing Browser does not fulfill the need of Visual weakened individual to get data from Internet. By focusing on the diverse articles the solution is provide such novel Software that will help to learn from the information from boundless wellspring of data that is Internet. It will give a complete help to such people by the usage of diverse utilities that is suggested with the solution very effectively. The various modules implemented with this system makes the solution distinct from other existing solutions .Integrating live Optical character recognition with maximum accuracy with pdf to audio file conversion module (TAF) is really exiting idea for to development of featured browser for visual impaired. And this integration of such various helpful modules is strongly implemented in this solution.

REFERENCES

- [1] D.Sasirekha, E.Chandra “TEXT TO SPEECH: A SIMPLE TUTORIAL” *International Journal of Soft Computing and Engineering (IJSCE)* Volume-2, Issue-1, March 2012
- [2] K.R. Aida-Zade, C. Ardil and A.M. Sharifova “The Main Principles of Text-to-Speech Synthesis System” *International Journal of Computer, Electrical, Automation, Control and Information Engineering* Vol:7, No:3, 2013.
- [3] Mahwash Ahmed, Shibli Nisar “Text-to-Speech Synthesis using Phoneme Concatenation” *International Journal of Scientific Engineering and Technology* Volume No.3 Issue No.2.
- [4] Md. Sipon Miah , and Tapan Kumar Godder “To Design Voice Control Keyboard System using Speech Application Programming Interface” *International Journal of Computer Science Issues*, Vol. 7, Issue 6, November 2010.
- [5] Ali Bahaloo “Increasing Accuracy of Speech Recognition Applications on Windows platform using .net framework” *University of Gothenburg Report no.2010.0128*.
- [6] Priyanka N.Waghela, Ankita A. Shirsekar, Sandhyarani D. Narayankar, Vijaya Pinjarkar “SPEECH ENABLED BROWSER USING MICROSOFT SPEECH API” *Proceedings of National Conference on New Horizons in IT - NCNHIT 2013*.
- [7] Md. Raihaan Kamarudin., Md. Aiman F. Md. Yusof.” *Low Cost Smart Home Automation via Microsoft Speech Recognition” International Journal of Engineering & Computer Science IJECS-IJENS* Vol:13 No:03
- [8] Bhushan Sonawane, Kiran Patil, Nikhil Pathak, Ram Gamane “Third Eye : An Image Explorer” *International Journal of Emerging Technology and Advanced Engineering* Volume 3, Issue 4, April 2013.
- [9] Itunuoluwa Isewon, Jelilii Oyelade, Olufunke Oladipupo “Design and Implementation of Text To Speech Conversion for Visually Impaired People” *International Journal of Applied Information Systems (IJAIS)* Volume 7– No. 2, April 2014.
- [10] Hao Shi and Alexander Maier “Speech-enabled windows application using Microsoft SAPI” *IJCSNS International Journal of Computer Science and Network Security*, VOL.6 No.9A, September 2006.
- [11] Parwinder pal Singh and Er. Bhupindersingh “Speech Recognition as Emerging Revolutionary Technology” *International Journal of Advanced Research in Computer Science and Software Engineering* Volume 2, Issue 10, October 2012.
- [12] Ayatullah Faruk Mollah, Nabamita Majumder, Subhadip Basu and Mita Nasipuri “Design of an Optical Character Recognition System for Camera- based Handheld Devices” *IJCSI International Journal of Computer Science Issues*, Vol. 8, Issue 4, No 1, July 2011
- [13] L.Santhoshi, K. Mahesh, Dr.K.Kuppusamy “Web Browsing For Visually Impaired People Through Handheld Devices “*International Journal of Engineering Research and Applications (IJERA)* Vol. 2, Issue 3, May-Jun 2012, pp.3024-3028