Human Machine Interface (Speech)

WAV: Voice Access to Web Information for Masses

Himanshu Chauhan, Pankaj Dhoolia, Ullas Nambiar, Ashish Verma
IBM Research -India, New Delhi {himchauh,pdhoolia,ubnambiar,vashish}@in.ibm.com

ABSTRACT– One of the main reasons for a large section of the world population to be left out of the internet revolution is, limited or no access to a computer due to economic, educational, cultural and age factors. Enabling masses to extract information from the web via voice will bring the Internet revolution to additional billions of people. In this paper, we describe a system called WAV (Web Access via Voice), that is a step in this direction. Departing from the traditional approaches of manually building a VoiceXML based site, the WAV system uses information from existing web sites to serve the user. Challenges to overcome include extracting contextually relevantinformation from the user and also from the pages returned by websites, reducing amount of information relayed to user over phone and maintaining the context of the conversation for easy renement based on feedback from the user. Our prototype system not only shows successful integration of many dierent technologies such as automatic speech recognition, scripts for web navigation, text to speech conversion, but also introduces a novel way of extracting information from web via voice in a programmatic manner. We describe initial solutions developed to tackle above challenges and demonstrate the feasibility of the system by describing proto type implementations on two popular web sites in India.

I. INTRODUCTION

The Internet and the world-wide-web, arguably the most important inventions of the last century, have changed the lives of billions of people in the world. Unfortunately, there are stillbillions of people who are unable to access the information on the world-wide-web. Any invention that allows the billions that are left out in the current Internet revolution to benet from the Internet will have a huge positive impact.

One of the main reasons for a large section of the world population to be left out of the internet revolution is, limited or no access to a computer/internet due to economic, educational, cultural and age factors. Although the penetration of Internet is low in the poor segment of the society, the use of telephones; particularly basic mobiles phones, is growing at an astounding rate. By November 2007, the total number of mobile phone subscriptions in the world had reached 3.3 billion, or half of the human population. Most of these subscriptions are in the developing regions of the world and for such people, the only easy-to-use medium for information retrieval is the basic mobile phone[1].

Web Access by Voice (WAV) is a system that focuses on decoupling the Information Content available on the World Wide Web from the Web Browsing methodology used to access it. It combines the ubiquity of the phone (both analog and mobile) with the information content available on the Web. WAV uses a novel methodology where the system performs web browsing instead of the user so that the user doesn't need to be familiar with the browsing and its nuances. The system works by taking the user's query, identifying the websites corresponding to user's query, gathering the required inputs from the user to extract the information, extracting the information related to user's query from web-sites, transforming it in a form consumable by the user over phone and supplying it to the user. Another important aspect of WAV is that the user doesn't need to know which websites have the information that the user wants. This really helps a large number of users who are not familiar with the Internet or the browsing concept and hence do not have the knowledge of various websites.

Challenges: The ultimate goal of providing access would be to build an intelligent natural language understanding system that parses a natural language query, understands the query, retrieves the information from the Internet and then extracts the answer for the query from the extracted information. With the current natural language parsers, such a system looks infeasible in the short term. However, speech recognition in a domain specic manner with a bounded number of words has been quite successful in spoken dialog systems [11]. WAV uses a carefully designed dialog manager to restrict the number of possible inputs by the user in a domain-constrained manner. There are still many other technical challenges:

[1] In this paper, all references to mobile phones are to basic mobile phones that do not give advanced capabilities like browsing Web pages or reading and responding to emails etc.

1. *Extracting Information for the User :* How can the system navigate the website to determine the information required by the user? Since many websitesprovide information through dynamic content generation, this step may require lling of forms on the user's behalf to extract the information. For example, suppose that the user is interested in nding out the schedule of trains from point A to point B. The user may not even be aware of the form that needs to be lled to get the schedule.

2 *Reducing Information for Voice-based Interaction* In many examples, the information returned from the website may be too much to read out to the user. Whereas a user can quickly choose the required piece of information in a visual medium, the voice medium necessitates that theinformation eitherbe summarized or be broken down into chunks so that the user is not given excessive information in a sequential manner.

3. *Rening the Query based on Previous Results* Just as in visual web browsers, the users typically rene or change their query based on the results obtained in the previous query. If no state is retained from the previousquery,the user wouldbeforced toprovide all thedetailsof thequery again. Byretaining thecontext fromthepreviousquery andproviding the missing arguments automatically the user is spared of repeating the information after each query.

In this paper, we introduce a website independent method of extracting information from the website. There are two aspects of this method. First, we use a language which de-couples the placement of form elements from the form element itself. The second aspect of our method is to keep a domain-specic translation table to allow the same method to work for multiple websites. For example, one travel websitemay use "OriginationPoint," and theotherwebsitemay use "From" as the listbox for the user to choose the destination. By keeping the web page prole, the user is insulated from such dierences between the websites.

II. SYSTEM ARCHITECTURE

As Figure 1 illustrates, WAV system has three main architectural components -The WAV conûguration studio, the WAV database, and the WAV engine. Rest of this section describes these architectural components, the usage scenarios they participate in, and the details of the comprising subsystems and components.

2.1 WAV Conguration Studio

While our primary focus in the WAV system is to address the needs of the masses, it is also important to make sure that the task of extending and conguring the reach of the WAV system doesn't become too daunting for a WAV service-provider that it starts to impact its protability. WAVConguration studiois a web-basedapplication thatis intended to address this. Powered by form-detector, result-detector, click-thruinteractionrecorder, and schemamatchers it assists a WAV system congurator by providing substantial automation in conguring the WAV system. A set of generators further assist in conguring new domains, by transforming detected set of forms, results, and interaction patterns to domain ontology, and domain conversation interface, based on congurator's training interactions with representative content-provider sites. The Studio populates the WAV database with the congurations, which are then picked by the WAV Engine to support end-users with corresponding conversations.

2.2 WAV Database

WAV database is the conguration repository where the artifacts of the WAV model described earlier, are stored. These artifacts are produced by a WAV service provider using the Conguration studio, and used by the WAV engine in its interactions with both the end-users and the specic WWW sites. Apart from the domain, and WWW-site congurations, WAV database also contains the actual user-interaction data. User Proler observes the patterns of interaction in the recorded data and derives personalized information and interaction reduction rules from it.

2.3 WAV Engine

WAV engine is the executional backbone of the WAV system. On one end it enables optimal voice conversations for the masses addressing the form-factor, while on the other, it integrates with the congured WWW-sites seamlessly to support those voice conversations.

2.3.1 Voice Interaction Engine

The Voice Interaction Engine component carries out the conversations with end-users. Domain focused voice congurations are leveraged by the engine to increase the performance of speech recognition. Further the learnt user and usage proles are leveraged to enhance the user-experience. As a voice driven user interface this engine has to provide three key functionalities : Recognition of User Speech, Relay of textual information in audio format, Management of conversation.

Automated Speech Recognition (ASR) is responsible for recognition of user speech. For WAV we use nite state speech grammars to limit what users can speak [5], thus improving recognition accuracy. Audio relay of textual informationis achievedby use of Text-to-speech (TTS) component. VoiceXML[12] format provides both ASR and TTS functionalities with the help of voice-browsers. WAV uses VoiceXML(VXML) dialogs tointeract withthe users.

Dialog Manager manages conversations between the system and users to provide better user experience. It stores caller specic details as User Prole and uses it for personalized menu choices andquick retrieval of userdata without prompting for them.

2.3.2 Information Retrieval Kernel

The Information Retrieval Kernel governs the conversations that the WAV system has with the congured WWW-sites. Receiving the user-query and inputs from the voice interaction engine, the information retrievalkernel transforms the query into corresponding site-specic queries, using the provider-prole congurations, and federates the query to the relevant sites. It then receives the site-specic responses, transforms them to the site independent domain conversation format and aggregates them, removing duplications. The Kernel then leverages the information reduction rules to optimize and sort the resulting response, and passes it to the voiceinteraction engine to communicate to the end-user.

III. WAV PROTOTYPE IMPLEMENTATION

We have implemented a prototype of WAV system based on the architecture presented above. The current implementation provides useful insights for building a scaled up system. At present the system does not allow free speech conversations and expects the user toprovideprecise answers to prompts. Most of the extracted information text is relayed asis on the web-page. The prototype incorporates information retrieval from web-sites of three dierent domains. For each domain, the conversation interface was created by observing label data (visible on the rendered page) for HTML inputs as captured in the CoScripter script. Static rules are used for reduction of information being relayed. At present these rules are manually congured based on domain knowledge, for example in travel domain ights are read out in increasing order of fares. Using techniques like context based information extraction[13] anddomain ontology an extraction schema for the domains retrieves relevant and concise information to be relayed back to the user.

Prototype for cheapest ight between a source-destination pair: Conguration process for extracting flight informationis started by creating a representative CoScripter script by manually performing a flight search. Figure 2 shows actual web-site form being filled and corresponding script, generated using Conguration Studio.

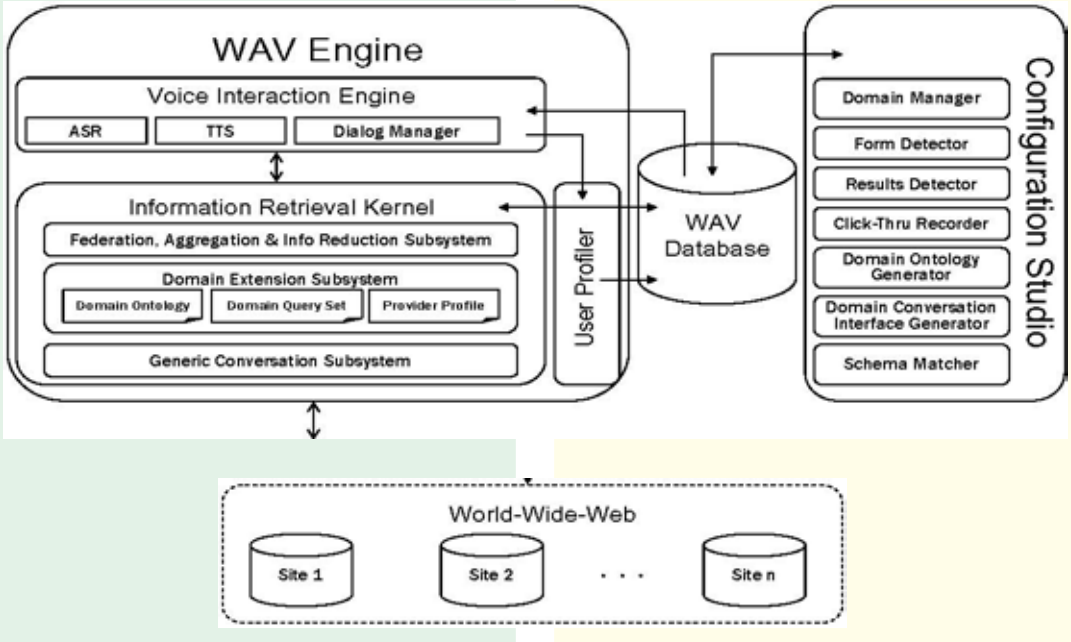

Figure 1: WAV System Architecture

On generation of the script, the domain ontology is augmented with label data from script such as 'From', 'To', 'One Way', 'Depart', 'Class' etc. Creation of Provider Prole follows the script creation step by using script to identify provider and inherit the URL. A common domain input schema is created using script statements and user provided input asparameters. These inputs are then mapped tocommon domain terms using inference models of domain ontology.

After completion of conguration step the domain is incorporated in the WAV system. Upon receiving a generic query for ights, it is translated to provider specic query format for each provider and then query Federation System federates specic queries for all the provider proles.

As seen in the gure, each row contains one ight detail. Extraction schema for the domain is congured to extract tuples of airline, departure time, fare and duration from each of these rows. These details are then collected and sorted based on the fare of each ight, putting cheapest ight on top, and relayed back to the user in audio format.

Prototype for train seat availability: Indian Railways provides information about all the trains schedules and availability of seats for reservation in trains through their website www.indianrail.gov.in. This prototype implementsatrain-domainconversation "Seat Availability Query", in a similar manner to that of ight details implementation. In this two step conversation, the conversation is initiated by providing source and destination stations along with the date of the travel to retrieve the list of trains available between the given source-destination pair on the particular date. In the second step of the conversation based on the train selection by the user, seat availability details are retrieved and communicated back.

IV. RELATED WORK

Speech interface to mobile web browsers has been providedprovided toperform simple web searches[9]. However this mode of internet access still requires familiarity with the web browsing concept and also it is dicult to use due to the tiny display available with the mobile phones. There also have been some eorts to perform web search through spokenquery[1]. Speech interface have also been developed for standard webbrowsers[10]. W3CVoice Browser Activity Group [5] has been working in this area and has come out with various standards, such as, VXML, SRGS, CCXML etc., mainly towards speech recognition, text-to-speech synthesis and natural language understanding aspects. However, there is not much work regarding how to perform the information extraction from dynamic web pages which are getting richer and more complex with usage of technologies, such as, Javascript and Ajax. There have also been some focussed work to improve the browsing eciency of visually impaired users starting from screenreaders [2] to more sophisticated approaches using content ltering and semi-automation[4,13]. In[4] an accessible interface is developed for CoScripter, a programming by demonstration solution, which helps the blind user perform apre-deneinternet task more eciently. In [13] the browsing time for a blind user

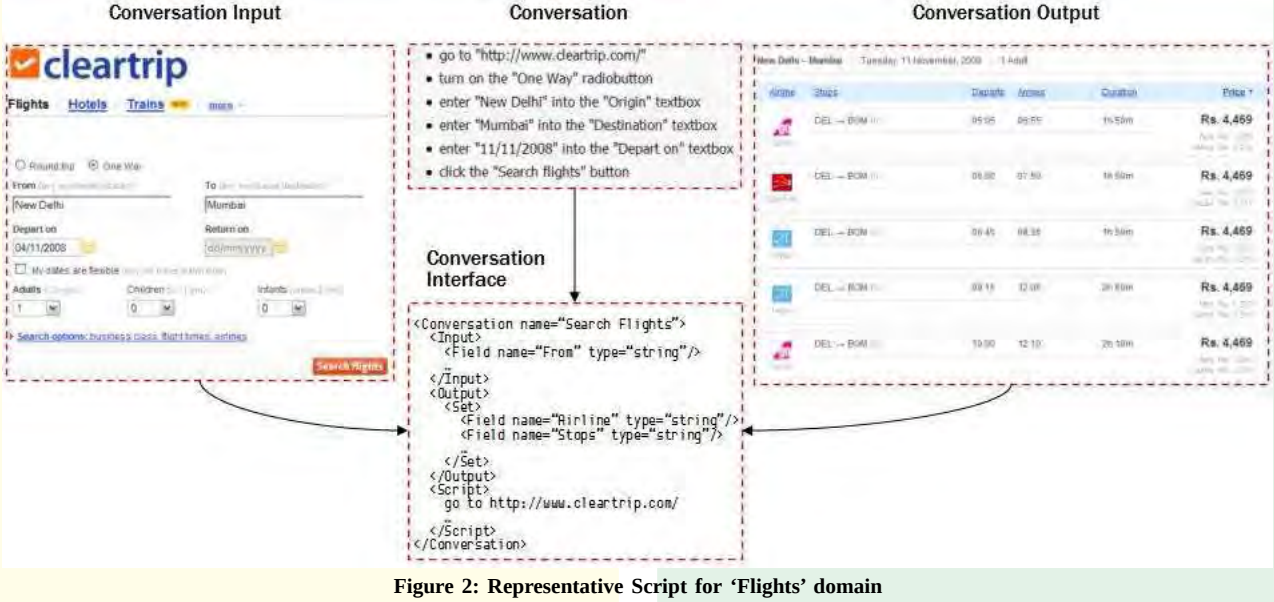

Figure 2: Representative Script for 'Flights' domain

is signicantly reduced as compared to using a normal screenreader by using smart content ltering mechanisms and content-aware web browsing concepts.

In a related eort at our lab, called WorldWide Telecom Web[7], aparallel worldwide webin the telecom domainis worked upon where the information would be stored completely in audio format, called voice sites, and would be browsed through a telecom browser[8]. The maindierence between SpokenWeb and WAV is that while in SpokenWeb the information resides in newly conceptualized voice sites, in WAV the information is extracted from the traditional worldwide websites.

V. CONCLUSION

In this paper, we present a system to access the information content on the Worldwide Web through voice for consumptionbypeople who eitherdon'thave access to computer or/and are not familiar to web browsing for various reasons. We described how the system extracts information related to a user's spoken query in a domain-dependent but website independent manner. This is performed by developing a domain ontology and a mapping between website independent terms and website dependent terms. We have described how designing such a system in a domain dependent manner helps in addressing the issues in information extraction from dierent websites. We have also shown how pseudo-naturallanguage script produced by CoScripter, can beusedfor such asystem. We described initial solutions developed and demonstrated the feasibility of the system by describing prototype implementations.

Currently, we are working on developing hierarchical domain structuring and ontology inheritance todevelop a scalable framework of WAV covering a large number of domains and hence a larger number of websites for information access.

6. REFERENCES

[1] Huixiang Gu, Jianming Li, Ben Walter and Eric Chang, "SpokenQuery forWebSearch and Navigation", Proc. of Internaltional WWW Conference, HongKong 2001

[2] http://www.freedomscientic.com

[3] N. Kushmerick, D. Weld, and R. Doorenbos. Wrapper Induction for Information Extraction. In Proceedings of the Fifteenth International Joint Conference on Articial Intelligence, pages 729–737. San Francisco, CA: Morgan Kaufmann, 1997.

[4] J.P. Bigham, T. A. Lau and J. W. Nichols,"TrailBlazer: Enabling Blind Users to Blaze TrailsThrough the Web", submittted to InternaltionalConference on Intelligent User Interfaces, Florida,2009.

[5] http://www.w3.org/voice

[6] G. Leshed, E. Haber, T. Matthews, T. Lau,"CoScripter: Automating and Sharing How-ToKnowledge in the Enterprise", CHI 2008, Florence,Italy, April 2008.

[7] Arun Kumar, Nitendra Rajput, Dipanjan Chakraborty, Sheetal K. Agarwal and Amit Anil Nanavati, "WWTW: The WorldWide Telecom Web", NSDR2007(SIGCOMM workshop), Kyoto, Japan 2007

[8] Sheetal Agarwal, Arun Kumar, Amit Anil Nanavati, Nitendra Rajput, "The WorldWide Telecom Web Browser", Proc. of International WWW Conference, Beijing, China, 2008

[9] http://mobile.yahoo.com/onesearch

[10] Dong Lin, Lin Bigin, Yuan Bao-Zong, "Using Chinese Spoken-Language Access to the WWW", Proc. of International Conference on WCCC-ICSP, Volume 2, pages:1321-1324, 2000

[11] Y. Gao, H. Erdogan, Y. Li, V. Goel and M. Picheny, "Recent Advances in Speech Recognition System for IBM DARPA Communicator", Proc. of EUROSPEECH, Denmark, 2001

[12] http://www.w3.org/TR/voicexml20/

[13] Jalal Mahmud, Yevgen Borodin and I.V. Ramakrishnan, "CSurf: A Context-Driven Non-Visual Web-Browser", International World Wide Web Conference WWW,(http://www.www2007.org)

Spoken Web: A Parallel WWW in Developing Regions

Sheetal K. Agarwal, Arun Kumar, Amit Anil Nanavati, Nitendra Rajput
IBM India Research Lab, 4, Block C, ISID Campus, Vasant Kunj, India

ABSTRACT— The Spoken Web is a voice-based equivalent of the World Wide Web (WWW), developed by IBM Research Laboratory, India, primarily designed for rural and semi-urban people to provide information of value to them through their mobile or landline phones. It will also help the government / industry / micro-business to reach out to rural population with their offerings and help the people at the Bottom of the Pyramid. The vision is to create an information ecosystem that helps provide Internet-like information services through phones.

MOTIVATION

The World Wide Web (WWW) has become a rich source of accessing information and services over the last decade. Hardly anyone reading this paper would have not accessed the WWW today, for accessing some kind of information. The Internet therefore is one of the most significant technologies that have changed our daily lives in the recent past. This has been made possible through the numerous information sources and applications available over the WWW. However the impact of theWWWis still not at the level of some basic facilities such as the railroad, electricity

Because of the low literacy rates, and the low Internet penetration in India, the PC based model to deliver information and services in the rural areas in India has not been as effective as in the western world.. There is a significant percentage of population that is still untouched by the WWW revolution and are either unaware of or are unable to catch the momentum. Even today, barely 17% of the world's population has access to the Internet. There are a variety of reasons that act as a hindrance for this technology to impact the remaining 83% section of the human population. Firstly, 53% of the world population lives below USD 2 per day – so they cannot afford a PC or high end phones and hence cannot access the Internet. Secondly, a significant portion of the remaining 30% is illiterate and semiliterate people who do not know how to operate a computer. Thirdly, most of the information and applications available on the Internet is hardly relevant to this section of the society. However, for this technology to become a commodity such as a road or electricity, a significant shift in paradigm is needed. Incremental improvements in terms of the sophisticated services, advanced user interfaces, easier application authoring techniques for the WWW do not appear to be helping in providing a shift of such a large scale. Interestingly, the telecommunication network does not face some of the challenges of the Internet world – from an acceptance perspective. The cost of a phone is significantly lower than a PC and, the learning required to operate a phone is negligible as compared to a PC, especially when the phone is used as a device to communicate in free speech.

Thus telecommunications have become a commodity for the common man and are a step closer towards achieving that status for the underprivileged as well. The Spoken Web, in our vision, has the potential to deliver to underprivileged, what WWW delivers to IT literate users today. In this paper, we present the concept of Spoken Web and illustrate a specific usage scenario to explain the working model. We also present the details of a specific service called VoiceSite to create services for the individuals.

INTRODUCTION

The Spoken Web [1] is a voice-based equivalent of the World Wide Web (WWW), developed by IBM Research Laboratory, India, primarily designed for rural and semi-urban people to provide information of value to them through their mobile or landline phones. It will also help the government / industry / micro-business to reach out to rural population with their offerings and help the people at the Bottom of the Pyramid. The vision is to create an information ecosystem that helps provide Internet-like information services through phones.

VoiceSites are voice driven applications that are created by the subscribers and hosted in the telecom network. They are represented by a unique phone number and can be accessed from any phone instrument,mobile or landline through an ordinary phone call to that number. The phone does not require any extra features or software to be installed on the device. VoiceSites are therefore analogous to websites in the WWW but can be accessed by dialing a phone number and information can be listened rather than being read or seen. Creation of a VoiceSite is made easy by the VoiGen system through which anyone can call up the VoiGen system and interact with it through voice. This can enable any illiterate person to create her VoiceSite. Such a system

enables easy local-content creation. All information in the VoiceSite is stored as audio messages that are recorded by making a phone call to the system. A blind person can create his/her VoiceSite by a simple phone interaction with the VoiGen system.

By answering simple questions as the ones shown in this interaction, a visually challenged user can easily create her VoiceSite. A person who needs a classical vocal teacher can access this VoiceSite and get her information. Such a simple mechanism can improve the business of the VoiceSite creator.Users can dial into Spoken Web through mobile phones or landline phones, *create* their own Voicesites (analogues to websites) in local languages, and *browse* voicesites created by others. An example in the rural setting is of a Voicesite Owner who gathers local information that has relevance to a specific cluster of villages, and uploads it onto his Voicesite. The villagers dial the Voicesite number and receive the information that they wants. All this is possible just by talking over the phone – in their own local language and dialect.

Spoken web has the ability to link multiple voicesites together [2]. User can call in to any of these voice sites and move from one to the other and the context get transferred seamlessly.

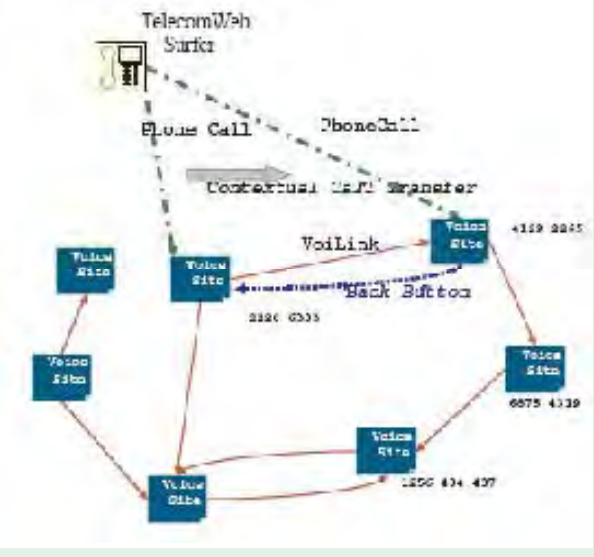

Figure 1. The Spoken Web

The Spoken Web has tremendous implications: For the visually impaired this means that they do not have to spend months learning how to use a computer and the various accessibility software such as screen readers. It also enables the non-PC literate people to access information and services that were hitherto unavailable to them through IT systems [3].

MOTIVATING SCENARIO

Sanjay has just moved to a house in a new city. He needs to fix several taps and electricity fittings. However since he is new in the locality, he does not know where to find the plumbers and electricians. He does not want to rely on the yellow pages since they have mostly stale limited information.

At the other end, Tom is a skilled electrician who works in the area where Sanjay lives. Tom has a lot of free time but is not able to get business since not many people know about him. Tom is an individual businessman who is not IT-savvy, so he cannot afford to advertise his skills through the Internet. However, since Tom works in different houses when he gets work, he carries a mobile phone so that people who know him can reach him for business.

To increase his business reach, Tom creates his VoiceSite. This can be done by making a phone call to the system and then specifying the different features about his service. When the system asks, Tom replies by mentioning his work area, charges, working hours and a list of customer references who can vouch for his good work. This information is used to create his VoiceSite. Sanjay makes a phone call to Tom's VoiceSite and gets all the details about Tom's work. The service charges and work times suit him. He then calls a couple of Tom's references to verify if he has done good work earlier. Satisfied with the responses, Sanjay finally asks Tom to provide the service to him.

The VoiceSite helps Tom in getting the business and it helps Sanjay in getting a right person for the services that he required.

We observe some salient capabilities of our system from the above scenario:

1 Using this system does not require to learn any new interaction modalities – just a voice conversation with the system is good enough. Thus semi-literate people create their VoiceSite through a voice-driven interface.

2 At no extra cost (other than the cost of making a phone call), the service providers and service requesters will be able to exchange the information through the VoiceSite.

3 It requires no extra device or hardware to create a voice avatar for a person. They can do this just by making calls through their mobile phones.

4 It makes the vendor accessible without having to establish a direct communication channel with the service providers.

A sample interaction of the service provider with the VoiceSite system is presented in Figure 2. The number of questions are made very few and the system asks the user to record her details. The system is not heavily dependent on the speech recognition errors since very few utterances of the end-user are converted to text – most is recorded. Moreover, since the generated VoiceSite of the service provider is in his/her own voice, callers can easily relate the VoiceSite to the actual person. When a service requester needs the service of a carpenter, it calls the VoiceSite of Ratan and interacts with the system as shown in Figure 4 to get information about the carpenter.

S: Welcome to the VOLAVATAR system. Please speak your welcome prompt
C: My name is Ratan and I work as a carpenter in south extension, malviya nagar and nearby areas
S: The welcome prompt has been recorded.
S: Say 1 if you wish to specify your references, Say 2 to specify the charges or Say 3 to specify your working hours.
C: One
S: Say the people and their contacts who can vouch for your work.
C: You can ask about me from Mr. Sanjay, his number is 9484012433. Or you can ask from Mr. Gaurav - 9844740948.
S: References have been recorded. What next?
C: Two
S: What are the charges that you ask for your service.
C: I charge $1 for an hour and $10 for a day. However the rates really depend on the type of job.
S: Charges have been saved. What next?
C: Three
S: What are your working hours.
C: I work from 9 am to about 8 pm - all seven days a week.
S: You have specified all the preferences. Should I save them?
C: Yes.
S: Your VOLAVATAR has been created. Thanks for using the system

Figure 2: A sample interaction of service provider with the Spoken Web system.

Both, the service provider and the service requester are able to provide and access information through the normal telephony network. The VoiceSite is able to provide the information about the mobile worker Ratan and thus can enhance his business prospects.

SPOKEN WEB DEPLOYMENTS

In all, there have been many applications deployed on the Spoken Web platform. In all the settings, end-users call up the VoiceSites by dialing a specific phone number from a regular phone. The VoiceSites are hosted on a centralized server. The VoiceSites are authored in VoiceXML and use a VoiceXML browser (Genesys Voice Platform, GVP, in our case) and a speech recognition server (Websphere Voice Server, WVS, in our case). Built on standard languages and supporting standard protocols, the Spoken Web VoiceSites themselves do no depend on the specific servers mentioned above.

The Spoken Web applications have been used by more than 10,000 users across several states in India [4,5]. In this section, we will provide a specific test deployment that highlights the ease of VoiceSite creation in the Spoken Web.

The specific application presented in this section, on receiving a call, carries out the following activities:

1 It uses custom recorded prompts (spoken in the local language) to educate the caller about the content he/she can put in his/her VoiceSite.

2 It prompts the caller to specify his/her preferences and records them.

3 It browses through the template and guides the creator to provide information to customize the template. E.g. a template of a mobile micro-businessman such as a plumber, allows the caller to provide basic information about him, followed by the ability to create reference/links to other users as well as professional information about his/her business.

4 On receiving all the inputs, it parses through the data obtained, and automatically generates a VoiceSite for the caller using a generation engine [4].

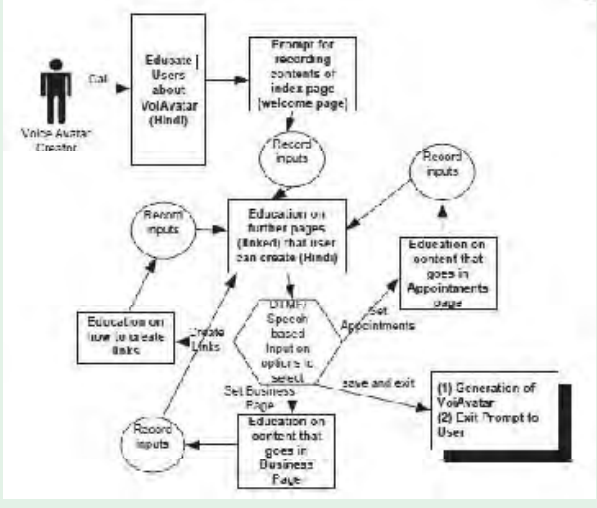

Figure 3: Control flow diagram of a voice template used to create the VoiceSite mentioned in this section.

Figure 3 shows the control flow diagram of a concrete template that we used in VoiceSite to enable small and micro businesses (plumbers, electricians, servants, home delivery services) in metropolitan cities to create their avatars. This voice template allows these classes of users to create their welcome page, create their business page where they provide information about their business, and also offers the ability to create a page to specify appointment hours. Apart from these, it allows these users to create links to other users (1) they

might want to connect to; (2) they use as references for their business.

The name of the author and affiliation should follow on separate lines in upper and lower case letters. Use the Times or Times Roman 10 point typeface.

SYSTEM EVALUATION

While the different Spoken Web applications have been vastly used by end-users, we present results from a specific study that indicates the ease of creation of VoiceSites for the abovecategory of users.

The subjects of this study were typically skilled labourers (such as electricians, plumbers, carpenters) who charge on the basis of the amount of work that is required of them. This profile is typically observed in developing countries where a decentralized and a disconnected set of labourers work independently. Most of them have had only about five to ten years of formal school education. Their yearly earnings are below Rs. 75000. They get business when households call them for work to fix things in their home/office. Their advertisement largely depends on the social network and is based on the word-of-mouth. Despite their low income, most of them carry a cell-phone since it helps in their business to be reachable (available on call as far as possible). They do not have a phone at their residence. We surveyed 12 subjects of which 3 were carpenters, 5 were plumbers, 3 were electricians and one was a drilling person.

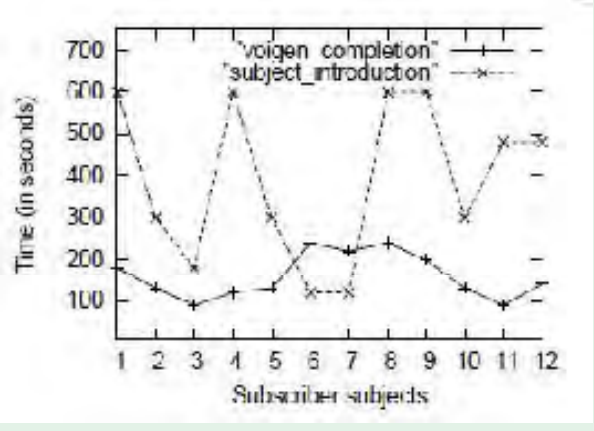

Figure 4: Time taken by each subject to understand the concept of Spoken Web and to build their VoiceSites.

Figure 4 shows the time that was spent to build the voice avatars for the 12 subjects. As seen, most subjects were able to generate their own voice avatars within 4 minutes. A 4 minute phone call in India costs less than 4 rupees. This is despite the fact that none of these subjects knew or had used an IVR before. Surprisingly most of them were comfortable in using the IVR. More importantly, all of them were able to identify the potential that a voice avatar can have in increasing their business. However each such phone call had to be preceded by an introduction to the concept which took about 5 minutes for each subject, as is seen in Figure 4. All showed tremendous interest in the concept of voice avatars and the fact that their work can be advertised without them actually requiring to purchase any additional equipment.

CONCLUSION

Spoken Web is an attempt to envision a service for the underprivileged communities, similar in theme to what WWW is to the IT literate users today. It enables masses to access information and services through voice driven channels. Information and services could be community created as well as leveraged from existing Internet infrastructure. We have summarized key technology enablers for this vision to be successful as well as attempted to articulate interesting research problems in this vision. Enabling voice-driven front-ends to websites and WWW services would only enable the underprivileged to access global information.

Spoken Web builds a vision of a service for users in developing regions that harnesses WWW services as well as the ones in the converged networks — under one umbrella. Further, it provides the means to create and sustain an ecosystem of local (and global) services, information and communities relevant to these underprivileged users.

REFERENCES

[1] A. Kumar, N. Rajput, D. Chakraborty, S. K. Agarwal, and A. A. Nanavati, "VOISERV: Creation and Delivery of Converged Services through Voice for Emerging Economies," IEEE WoWMoM, Helsinki, Finland, June 2007.

[2] S. K. Agarwal, D. Chakraborty, A. Kumar, A. A. Nanavati, N. Rajput, "HSTP: Hyperspeech Transfer Protocol," *ACM Hypertext*, Manchester, UK, 10-12 September 2007.

[3] A. Kumar, N. Rajput, S. K Agarwal, D. Chakraborty, A. A. Nanavati, "Organizing the Unorganized Employing IT to Empower the Under-privileged," *WWW*, April 2008.

[4] S, Agarwal, A. Kumar, A. Nanavati, N. Rajput, "Content Creation and Dissemination by-and-for Users in Rural Areas," *ICTD* 2009.

[5] N. Patel, D. Chittamuru, A. Jain, P. Dave, and T. S. Parikh, "Avaaj Otalo ? A Field Study of an Interactive Voice Forum for Small Farmers in Rural India," *CHI* 2010.

Security Issues in Human Machine Interface for Web

Pooja Dhiman*, Suman Kathpalia**
Lecturer, Chitkara Institute of Engineering & Technology(Jhansla), Punjab
** poojadhiman23@gmail.com,**sumankathpalia@gmail.com*

ABSTRACT– Unlike traditional systems, it's easy to see why web-based systems are gaining popularity. Web-based systems install and run client applications from any web-browser and when users login they always get the most recent version of an application. There are no client licenses manage, no tedious software installations, no application files to copy over and no communication configurations to setup. IT departments are willing to embrace technology they understand. The bottom line is, web-based HMIs systems fit well with the rest of the enterprise and facilitate the smooth flow of information throughout an organization without unnecessary difficulty and expense. When potential users first consider using web-based technology they usually ask about security. Just how secure are web-based systems? HMI and SCADA security "one of the most serious risks to our national security". Web-based systems, on the other hand, are already positioned to leverage standard and proven web security techniques as administered by IT departments. The Login scheme also has support and maintenance issues. These panels frequently are a source of a single point of failure. If the panel fails, operators can no longer control the process.

KEYWORDS: HMI, RTUs, PLCs, Inductive Automation

INTRODUCTION

Introduction: A Human-Machine Interface or HMI is the apparatus which presents process data to a human operator, and through this, the human operator monitors and controls the process.

- A supervisory (computer) system, gathering (acquiring) data on the process and sending commands (control) to the process.
- Remote Terminal Units (RTUs) connecting to sensors in the process, converting sensor signals to digital data and sending digital data to the supervisory system.
- Programmable Logic Controller (PLCs) used as field devices because they are more economical, versatile, flexible, and configurable than special-purpose RTUs.
- Communication infrastructure connecting the supervisory system to the Remote Terminal Units.

Why it is required:- Once considered impractical for applications requiring responsive animation and real-time control, a new breed of web-based HMI system is starting to appear on plant floors and in manufacturing enterprises. "Java (web) based systems can now deliver sub-second response, rich animation and natural integration with other parts of the corporate information infrastructure." Unlike traditional systems, these web-based systems can economically be extended to every aspect of a business such as QC, maintenance, logistics, plant manager, and so forth. Now every participant in the manufacturing cycle can have unprecedented access to vital plant production information.

It's easy to see why web-based systems are gaining popularity. Web-based systems install and run client applications from any web-browser and when users login they always get the most recent version of an application. There are no client licenses manage, no tedious software installations, no application files to copy over and no communication configurations to setup. IT departments are willing to embrace technology they understand. All this is in sharp contrast to traditional systems. The economic advantages of using web-based systems are compelling. The bottom line is, web-based HMIs systems fit well with the rest of the enterprise and facilitate the smooth flow of information throughout an organization without unnecessary difficulty and expense.

SCADA:- The term SCADA usually refers to centralized systems which monitor and control entire sites, or complexes of systems spread out over large areas (anything between an industrial plant and a country). Most control actions are performed automatically by Remote Terminal Units ("RTUs") or by programmable logic controllers ("PLCs"). Host control functions are usually restricted to basic overriding or *supervisory* level intervention. For example, a PLC may control the flow of cooling water through part of an industrial process, but the SCADA system may allow operators to change the set points for the flow, and enable alarm conditions, such as loss of flow and high temperature, to be displayed and recorded. The feedback control loop passes through the RTU or PLC, while the SCADA system monitors the overall performance of the loop.

Data acquisition begins at the RTU or PLC level and includes meter readings and equipment status reports that are communicated to SCADA as required. Data is then compiled and formatted in such a way that a control room operator using the HMI can make supervisory decisions to adjust or override normal RTU (PLC) controls. Data may also be fed to a Historian, often built on a commodity Database Management System, to allow trending and other analytical auditing.

SCADA systems typically implement a distributed database, commonly referred to as a *tag database*, which contains data elements called *tags* or *points*. A point represents a single input or output value monitored or controlled by the system. Points can be either "hard" or "soft". A hard point represents an actual input or output within the system, while a soft point results from logic and math operations applied to other points. (Most implementations conceptually remove the distinction by making every property a "soft" point expression, which may, in the simplest case, equal a single hard point.) Points are normally stored as value-timestamp pairs: a value, and the timestamp when it was recorded or calculated. A series of value-timestamp pairs gives the history of that point. It's also common to store additional metadata with tags, such as the path to a field device or PLC register, design time comments, and alarm information.

Applications projects :-

1. VOICE - A Voice Oriented Interactive Computing Environment
2. OCR And Speech Recognition For Oriya Language
3. Adapting Question Answering Techniques to TheWeb
4. Foresight - A Personalised Syntax Based Prediction System for Natural Language Text Input into Computers
5. NLP Research In The CSR Lab Of Tata InfoTech
6. A Hybrid Scheme For Hand printed Numeral Recognition Based On A Self-Organizing Network and MLP-Based Classifiers
7. On Developing High Accuracy OCR Systems for Telugu and Other Indian Scripts
8. Recognition of Hand printed Bangla Numerals Using Neural Network Models
9. Self-Organizing Neural Network-Based System for Recognition of Handprinted Bangla Numerals

SECURITY PRINCIPLES

Introduction

To aid in designing a secure information system, NIST compiled a set of engineering principles for system security. These principles provide a foundation upon which a more consistent and structured approach to the design, development, and implementation of IT security capabilities can be constructed. While the primary focus of these principles is the implementation of technical controls, these principles highlight the fact that, to be effective, a system security design should also consider non-technical issues, such as policy, operational procedures, and user education.

The principles described here do not apply to all systems at all times. Yet each principle should be carefully considered throughout the life-cycle of every system. Moreover, because of the constantly changing information system security environment, the principles identified are not considered to be an inclusive list. Instead, this document is an attempt to present in a logical fashion fundamental security principles that can be used in today's operational environments. As technology improves and security techniques are refined, additions, deletions, and refinement of these security principles will be required. Each principle has two components. The first is a table that indicates where the principle should be applied during the system life-cycle. The second is an explanatory narrative further amplifying the principle.

The five life-cycle planning phases used are defined are:

- Initiation Phase
- Development/Acquisition Phase
- Implementation Phase
- Operation/Maintenance Phase
- Disposal Phase.

System Life-Cycle Description

The following brief descriptions of each of the five phases of the system life-cycle are

Initiation: During the initiation phase, the need for a system is expressed and the purpose of the system is documented. Activities include conducting a sensitivity assessment.

Development/Acquisition: During this phase, the system is designed, purchased, programmed, developed, or otherwise constructed. This phase often consists of

other defined cycles, such as the system development cycle or the acquisition cycle. Activities include determining security requirements, incorporating security requirements into specifications, and obtaining the system.

Implementation: During implementation, the system is tested and installed or fielded. Activities include installing/turning on controls, security testing, and accreditation.

Operation/Maintenance: During this phase, the system performs its work. The system is almost always being continuously modified by the addition of hardware and software and by numerous other events. Activities include security operations and administration, operational assurance, and audits and monitoring.

Disposal: The disposal phase of the IT system life-cycle involves the disposition of information, hardware, and software. Activities include moving, archiving, discarding or destroying information and sanitizing the media

Security Issues

When potential users first consider using web-based technology they usually ask about security. Just how secure are web-based systems? The question is especially valid now that post 9/11 committees have deemed HMI and SCADA security "one of the most serious risks to our national security." Traditional vendors rely heavily on "security by obfuscation" which has never been considered a safe practice. Web-based systems, on the other hand, are already positioned to leverage standard and proven web security techniques as administered by IT departments.

It's only be a matter of time before legislation mandating minimum HMI and SCADA security requirements will surface. Traditional providers will likely have to overhaul their products to come into compliance. They will welcome this day since they will sell lots of mandated security upgrades

Component vulnerabilities within an HMI system

To minimize existing security gaps, companies need to first understand where potential vulnerabilities typically lie within the system. Powerful software features, along with the advancements in automation hardware and industrial communications, have made control systems multi-layered, complex and susceptible to threats. An HMI/SCADA system level of security is best understood if broken down into two major elements: Communication and Software Technology.

Communication

Communication advancements have made large-scale HMI/SCADA system implementations successful for many industry applications. There are two levels of communication that exist within the system—information technology (IT) and the field, which have notable security level differences.

Software technology

Software over the years has largely become feature-bloated as companies keep adding new capabilities while maintaining all of the existing ones, increasing the complexity of software security. There are two separate but dependent software technologies in the system, the HMI/SCADA software and the Platform Operating System, which have distinct differences when it comes to security.

HMI Software - Most HMI/SCADA software installations have either external network connections or direct Internet-based connectivity to perform remote maintenance functions and/or connect up to enterprise systems. While these types of connections help companies reduce labor costs and increase the efficiency of their field technicians, it is a key entry point for anyone attempting to access with a malicious intent.

Some vulnerability is minimized by the nature of system design and HMI/SCADA software design, whereby the fundamental principles and canons of engineering mandate safe and reliable systems. This ensures a basic level of security to protect against an intruder. Engineers design systems with intentionally broken auto-mated chains—meaning in some cases functions require physical confirmation prior to the software performing commands and in other cases; the SCADA software only does a portion of the command, requiring one or many additional manual steps to execute the function. Inherent system security is best surmised at the software and hardware levels.

Visualization tool that provides a means for dynamic operator input and visualization as a flexible information terminal, the reality is that HMI/SCADA software capabilities are much more exhaustive. When elements are added such as control and logic capabilities, system engineers must examine the risk from a potential failure standpoint and the extent of control that is allowed without being in line of sight of the area being controlled.

Hardware

Techniques to ensure safe control, either physically or by the HMI/SCADA software. Thousands of

individual devices and RTUs can exist in a system and are typically implemented with an area-based manual or automatic control selection; field technicians use manual control to perform maintenance or to address a software failure—locking out the software control and establishing local control.

Functionally speaking, HMIs haven't changed much over the past five years. "HMIs that just do operator interface tasks are a commodity, and you can buy them dirt cheap off the Internet…The real action is in HMIs that provide web access, interface to higher-level enterprise software, perform MES functions".

Seeing What's Next

Applying some predictive theories to this industry suggests incumbent HMI vendors will continue to service their large existing market without much change. They will probably not compete with their own model. On the other hand, web-based vendors will find success selling where traditional vendors have failed; to those companies who refuse to spend big bucks on systems perceived as being unnecessarily complex, cumbersome and overshooting needs. This is likely to lead to explosive growth for web-based systems in market segments which have been unfulfilled by traditional systems.

Anyone familiar with manufacturing knows the majority of factories barely implement information technology at the plant floor level. There are exceptions, but when you see clipboards being used to record schedules, downtime and production, when you envision how things should be done, you finally come to realize this is a vast untapped market.

There is an accelerating pace of web-based systems being installed in what was essentially a non-consuming market. Users are finally getting what they want – the functionality of an HMI with the economics of a web browser. The real question is not whether web based control systems are an emerging trend – they cannot be stopped, but rather which vendors are poised to jump on the bandwagon and deliver the technology.

CONCLUSION

That's why it's important for companies to better understand the vulnerabilities of HMI/SCADA systems pose a serious where vulnerabilities exist within their systems and to take threat, and the complexity of multi-layered technologies a proactive approach to address those susceptible areas make it difficult to completely secure one's operation. As off-the-shelf HMI/SCADA vendors offer software solutions discussed in this paper, the inherent safe design of most with security-based capabilities, which can help companies HMI/SCADA systems offers some protection, but they are by enhance the protection of their critical infrastructure assets no means enough to fully protect systems and reduce costs for a sustainable competitive advantage.

REFERENCES:

www.automation.com/.../hmi.../web-based-hmi-an-emerging-trend

http://en.wikipedia.org/wiki/User_interface

http://www.researchandmarkets.com/reports/358954

http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6V0N-4VF56PK-5&_user=10&_coverDate=04%2F30%2F2009&_rdoc=1&_fmt=high&_orig=search&_sort=d&_docanchor=&view=c&_searchStrId=130 7822603&_rerunOrigin=google&_acct=C 000050221&_version=1&_urlV ersion=0&_userid=10&md5=c877bf bb917ee0bf9a92e657b99adfa4

http://pricerscorporation.com/index.php?option=com_content&view=article&id=44:human-machine-interface&Itemid=69

http://www.topbits.com/scada.html

http://www.aof.mod.uk/aofcontent/tactical/hfi/content/hfi_hmi.htm

http://www1.elsevier.com/homepage/saf/infosecurity/research/1206_scada.pdf

http://en.wikipedia.org/wiki/User_interface

http://www.indusoft.com.br/article.php?type=article&artid=30&lan=en/nternet/intranet/HMI/SCADA/FTP/HTML/XML/SOAP/UDDI/bandwidth/VPN/integrity/authenticity/confidentiali

http://www.sciencedirect.com/science?_ob=ArticleURL&_udi=B6TYV-4S7SV4F-5&_user=10&_coverDate=03%2F31%2F2009&_rdoc=1&_fmt=high&_orig=search&_sort=d&_docanchor=&view=c&_searchStrId=1307918617&_rerunOrigin=google&_acct=C000050221&_version= 1&_urlVersion=0&_userid=10&md5=85a3be130aa4af9b292c93681c7db03a

http://www.simulation.com/Corporate/multimedia/publications/pdf/hmi_generation.pdf

http://www.automationworld.com/blog-6518

http://blog.rootshell.be/2010/03/06/scada-from-a-security-point-of-view/

http://www.isa.org/filestore/Division_TechPapers/GlassCeramics/TP04AUTOW046.pdf

http://tdil.mit.gov.in/IntegratedEnvironment.pdf

http://www.wisegeek.com/what-is-human-machine-interface.htm

http://www.actel.com/products/solutions/hmi/default.aspx

http://www.tascomp.com/index.php/article/articleview/213/1/6/

Accessibility of all Information to all people

Veera Raghavendra and Kishore Prahallad
International Institute of Information Technology, Hyderabad.
raghavendra@iiit.ac.in kishore@iiit.ac.in

ABSTRACT— Our paper focus on issues involved in accessibility of information of types: speech, text, to all people which include literate, illiterate, visually challenged and differently abled persons. We discuss the issues involved, and provide suggestions or recommendations of required standards to be developed for various language technology components involved in human-machine interaction over web. We believe many such standards are not available as on date, and development of such standards aid in building better interfaces both on desktops as well as on mobile to provide accessibility of all information to all people.

KEYWORDS— Speech-to-Speech, ASR, TTS, Machine Translation, Information retrieval.

I. INTRODUCTION

Now-a-days people are highly depending web for their day-to-day life. The users of the web are literate, illiterate and visually challenged people. Literate users can manage the web without any issues. But, illiterate and visually challenged users require a literate help. The goal of the web is to facilitate communication between people who speak different languages and due to the increasingly globalizing world economy, humanitarian services and national security, there is an ever increasing demand for speech-to-speech translation. While substantial progress has been made over the past decades in each of the related areas of Automatic Speech Recognition (ASR), Machine Translation (MT), Information Retrieval (IR), Summarization, Natural Language Processing (NLP), Text-to-Speech (TTS) system. W3C is so far created standards in the areas of ASR and TTS. More standards are required for utilizing IR, MT and summarization and more over global standards are required to utilize them altogether.

The rest of the paper is organized as follows. Section 2 describes the each component elaborately. In Section 3, available W3C standards for various components are discussed. Section 4 discusses what standards are required for each component.

II. COMPONENTS OF SPEECH-TO-SPEECH

Automatic Speech Recognition: Speech recognition is the process of converting a speech signal, captured by a microphone or a telephone, to a set of words. Speech recognition system can be characterized by many parameters. An isolated-word speech recognition system requires that the speaker pause briefly between words, whereas a continuous speech recognition system does not. Spontaneous, or extemporaneously generated, speech contains disfluencies, and is much more difficult to recognize than speech read from script. The conventional statistical framework employed to accomplish the speech recognition comprises three major components – acoustic models, language model, and the pronunciation dictionary. Speech recognition applications include voice dialling, call routing, content-based spoken audio search, simple data entry, and medical transcription.

Machine Translation: Machine translation is a automated process of converting the text from one language to another. For example to Telugu text to Hindi text or English text to Hindi text. To process any translation, human or automated, the meaning of a text in the original language must be fully restored in the target language. Translation is not a mere word-to-word substitution. A translator must interpret and analyse all of the elements in the text and know how each word may influence another. This requires extensive expertise in grammar, syntax (sentence structure), semantics (meanings), etc., in the source and target languages, as well as familiarity with each local region. Current machine translation software often allows for customization by domain or profession (such as weather reports) — improving output by limiting the scope of allowable substitutions. This technique is particularly effective in domains where formal or formulaic language is used. Human and machine translation each have their share of challenges. For example, no two individual translators can produce identical translations of the same text in the same language pair, and it may take several rounds of revisions to meet customer satisfaction. But the greater challenge lies in how machine translation can produce publishable quality translations.

Information Retrieval: The techniques of storing and recovering and often disseminating recorded data especially through the use of a computerized system. Automated information retrieval systems are used to reduce what has been called “information overload”. Many universities and public libraries use IR systems to provide access to books, journals and other documents.

Web search engines are the most visible IR applications. Two main approaches are matching words in the query against the database index (keyword searching) and traversing the database using hypertext or hypermedia links. Keyword searching has been the dominant approach to text retrieval since the early 1960s; hypertext has so far been confined largely to personal or corporate information-retrieval applications.

Summarization: Summarization is the restating of the main ideas of the text in as few words as possible or in a new, yet efficient, manner. There are different types of summaries depending what the summarization program focuses on to make the summary of the text, for example generic summaries or query relevant summaries (sometimes called query-biased summaries). Summarization systems are able to create both query relevant text summaries and generic machine-generated summaries depending on what the user needs. Summarization of multimedia documents, e.g. speech, pictures or movies are also possible. Some systems will generate a summary based on a single source document, while others can use multiple source documents (for example, a cluster of news stories on the same topic). These systems are known as multi-document summarization systems.

Text-to-Speech System: A Text-to-speech system deals with conversion of text into spoken form. Now-a-days, TTS systems are used in many applications such as car navigation systems, information retrieval over telephone, voice mail, language education, screen readers, speech-to-speech translation systems and so on. The goal of a TTS system is to synthesize speech with natural human voice characteristics and, furthermore, with various speaker specific individualities and emotions. TTS system comprise of mainly two components; text analysis and waveform generation. Text analysis includes dividing the text into sentences and words, assigning syntactic categories to words, grouping the words within a sentence into phrases, identifying and expanding abbreviations, recognizing and analysing expressions such as dates, fractions, money, and grapheme-to-phone conversion. The second component is generally referred to as a synthesizer which generates the speech waveform for the given sequence of phones.

III. W3C STANDARDS FOR VARIOUS COMPONENTS

Speech Synthesis Mark-up Language (SSML): SSML [1] is responsible for rendering a document as spoken output and for using the information contained in the mark-up to render the document as needed by author. The following are the six major processing steps undertaken by a synthesis processor to convert marked-up text input into automatically generated voice output.

1. XML parse: An XML parser is used to extract the document tree and content from the incoming text document. The structure, tags and attributes obtained in this step influence each of the following steps. A simple English example is “cup<break/>board”; the synthesis processor will treat this as the two words “cup” and “board” rather than as one word with a pause in the middle.

2. Structure analysis: The structure of a document influences the way in which a document should be read. For example, there are common speaking patterns *associated with paragraphs and sentences.*

Mark-up support: The <p> and <s> elements defined in SSML explicitly indicate document structures that affect the speech output. A <p> element represents a paragraph. An <s> element represents a sentence.

```
<p>
<s>This is the first sentence of the paragraph.</s>
<s>Here’s another sentence.</s>
</p>
```

3. Text normalization: The written text may contain non-standard words. Such as numbers, dates, telephone numbers, etc., Text normalization is an automated process of the synthesis processor that performs this conversion. For example, for English, when “$200” appears in a document it may be spoken as “two hundred dollars”.

Mark-up support: The **say-as** element can be used in the input document to explicitly indicate the presence and type of these constructs and to resolve ambiguities. many acronyms and abbreviations can be handled by the author via direct text replacement or by use of the **sub** element, e.g.W3C can be written as World Wide Web Consortium.

```
<sub alias=”World Wide Web Consortium”>W3C</
sub>
```

4. Text-to-phoneme conversion: Word pronunciations may be conveniently described as sequences of phonemes, which are units of sound in a language that serve to distinguish one word from another.

Mark-up support: The **phoneme** element allows a phonemic sequence to be provided for any word or word sequence.

```
<phoneme alphabet="ipa" ph="t&#x259;mei&#x325;&#x27E;ou&#x325;"> tomato </phoneme>
```

5. Prosody analysis: Prosody is the set of features of speech output that includes the pitch (also called intonation or melody), the timing (or rhythm), the pausing, the speaking rate, the emphasis on words and many other features.

Mark-up support: The **emphasis** element, **break** element and **prosody** element may all be used by document creators to guide the synthesis processor in generating appropriate prosodic features in the speech output.

```
That is a <emphasis level="strong"> huge </emphasis>
```

6. Waveform generation: There are many approaches to this processing step so there may be considerable processor-specific variation.

Markup support: The **voice** element allows the document creator to request a particular voice or specific voice qualities (e.g. a young male voice). The audio element allows for insertion of recorded audio data into the output stream.

```
</voice>
<voice name="Mike">I want to be like Mike.</voice>
```

VoiceXML: VoiceXML [2] is an XML language for writing Web pages you interact with by listening to spoken prompts and jingles, and control by means of spoken input. VoiceXML brings the Web to telephones. VoiceXML has been carefully designed to give authors full control over the *spoken dialog* between the user and the application. The application and user take it in turns to speak: the application prompts the user, and the user in turn responds.

VoiceXML documents describe:

- spoken prompts (synthetic speech)
- output of audio files and streams
- recognition of spoken words and phrases
- recognition of touch tone (DTMF) key presses
- recording of spoken input
- control of dialog flow
- telephony control

The following example offers a menu of three choices: sports, weather or news.

```
<?xml version="1.0"?>
<vxml version="2.0">
<menu>
 <prompt>
 Say one of: <enumerate/>
 </prompt>
 <choice next="http://www.sports.example/start.vxml">Sports </choice>
 <choice next="http://www.weather.example/intro.vxml">Weather</choice>
 <choice next="http://www.news.example/news.vxml">News</choice>
 <noinput>Please say one of <enumerate/></noinput>
</menu>
</vxml>
```

This dialog might proceed as follows:

Computer:	Say one of: Sports; Weather; News.
Human:	Astrology
Computer:	I did not understand what you said. (a platform-specific default message.)
Computer:	Say one of: Sports; Weather; News.
Human:	Sports
Computer:	(proceeds to http://www.sports.example/start.vxml)

VoiceXML integrated speech synthesis, speech recognition and telephone interface together. Similarly, all other components of the speech-to-speech system need to be established and combined together.

IV. REQUIRED STANDARDS FOR EACH COMPONENT

So far W3C created standards for speech synthesis and speech recognition. To achieve speech-to-speech communication over web, W3C have to come up with some standards for other components. Following are the few required standards for each component.

Text-to-Speech System: This component plays major role in providing accessibility to visually challenged people. Now-a-days many news papers are putting their content over web. But, these vendors use their own fonts and follow their own standards for creating web page. Instead W3C should create standards to these news websites. Few standards are as follows:

1. Font type Vs Unicode: Unfortunately, there exists several popular websites which provide news in their local fonts instead of in Unicode.

2. Structure: Separate tags for headings, short stories, full stories, headlines. Such structure would help in accessing the data for the purpose of TTS, MT and summarization.

Machine Translation: A new and great initiation required for creating machine translation standards similar to SSML or VoiceXML. The standards should address the MT aspects such as.

1. What is the source language?

2. What is the target language?

3. POS information of source language text.

4. Possible word-to-word mappings (In case of foreign words).

Another issue is the integration of MT with TTS and ASR. So far, the integration or APIs to integrate MT, TTS and ASR is specific to a particular implementation party.

Information Retrieval and Summarization: Information retrieval and summarization are useful during speech-to-speech communication. If the user is looking for cross-information retrieval from the web, the system has to convert the text into target language and corresponding information has to be retrieved. To achieve this W3C has to provide some standards. Following are the few standards.

1. Source language key words need to be searched.
2. What kind of documents need to be searched.

V. DISCUSSION AND CONCLUSION:

An important aspect to be considered in further defining the standards is the user type. So far, the standards seem to be mostly applicable to common persons. However, with the role of ICT beings enhancing, differently abled persons, especially visually challenged are accessing the information over WWW. There are a few commercial and academically developed screen reading software which provide accessibility to the information. However, there seems to no standards either in their interfaces or in the way the information is available on the WWW for these screen reading software. For example, a user after getting acquaintance with a particular screen reading software may find difficulty in switching to other screen reading software as the interfaces (keys performing specific function) might differ, Also from the developers perspective, the future of information access lies in multilingual speech-to-speech mode, where the role of MT, IR, summarization play a roles apart from ASR and TTS engines. Sufficient and efficient standards need to be evolved in order to integrate these various components for providing access of information to all people. Yet another direction which is not discussed fully in this paper is multi-modal aspect of multi-lingual speech-to-speech access of information. It would not be difficult to image to use both audio and video in the near future to access information. Hence standards may to have generic to incorporate video mode of information access too.

REFERENCES

[1] Speech Synthesis Mark-up Language, http://www.w3.org/TR/speech-synthesis/.

[2] VoiceXML, http://www.w3.org/Voice/.