

The Web of Voices: how to connect 4.5 billion internet-less people to the Web

Chris van Aart¹, Anna Bon¹, Hans Akkermans¹, Victor de Boer¹,
Stephane Boyera², Wendelien Tuyp³, Nana Gyan¹

¹ The Network Institute, VU University Amsterdam, the Netherlands

² World Wide Web Foundation

³ VU -Centre for International Cooperation, VU University Amsterdam, the Netherlands

Abstract. The foremost World Wide Web challenge is to connect the 4.5 billion now internet-less people to the Web. A lot of internet-less people live in developing countries where local conditions pose enormous challenges on the design of new interfaces to the Web. We envision the Web of Voices as primary interface to knowledge on the Web. Enabling mobile voice dialogues with knowledge on the Web raises a number of interesting research challenges (1) Voice based access to and management of information and (2) Web of Conversations enabling application of web based knowledge via speech. These challenges will bring new types of multidisciplinary research and technological spin-off that will step far beyond the currently keyword, tagging, hyperlink-based and semantic web. These outcomes will turn the web from a passive knowledge rich-environment accessible for few into an active supporting personalized environment that is accessible for all.

1 Introduction

The foremost World Wide Web challenge is to connect the 4.5 billion now internet-less people to the Web. The World Wide Web Foundation envisages a world in which every person can have access to the Web. A lot of internet-less people live in developing countries, where fortunately mobile telephony is highly available [1]. Mobile cellular penetration in the developing world has reached 70% at the end of 2010 just 6 years after reaching 70% in the developed world [2]. This wide availability of mobile phones creates opportunities for economic innovation based on Web technology. However, given the oral tradition and the illiteracy of many in African countries there are challenges on the design of new interfaces to the Web[3].

We envision a Web of Voices as interface to knowledge on the Web. Voice interfaces are not new - we are aware of fictional talking robots and androids such as Hal, Data, Marvin [4] and of existing software such as personal and interface agents. Recently we have even seen Watson [5] triumphing over human players on the tv quiz-show Jeopardy! One of the visions of the Semantic Web is having software agents working for us on the Web [6]. Yet, Watson and other

“intelligent” entities are still not fully capable of performing vocal assistance with more daily tasks such as advise on agriculture, teaching new tasks, inventing concepts nor motivating people.

A lot of knowledge is already available on the Web in machine-readable forms. Access to this knowledge is facilitated by web services that are device driven, e.g. one has to create a query by typing, to click with a mouse pointer or to work with fingers to order results. The availability of mobile telephones can aid to the promise of World Wide Web availability. Enabling vocal access and management of knowledge on the Web raises a number of interesting research challenges (1) Voice based information management and (2) Voice based knowledge usage. For every challenge we present possible usages and research directions.

2 Challenge 1: Voice based information management

Ten years ago hardly anyone had foreseen the impact of mobile telephony and software, world-wide. Almost half a million mobile apps are available for a range of mobile platforms. Most of these apps have a graphical user interface and some form of connection to the Web. The dominant knowledge carriers of the current Web are text based. We made a lot of advances adding semantic infrastructures on top of the text-driven Web. However, the interfacing to these infrastructures are mainly keyword and text-driven. The text on the Web carries style and form of the dominant languages it is written in. At the same time, research into voice based access to the (semantic) web has largely focused on widely spoken languages (English, Chinese)

Speech is the most natural way for humans to communicate and exchange information, knowledge, ideas and feelings. So, rather bringing the Web to a voice device, such as the mobile telephone, we should bring voices to the Web. Many voice- driven systems such as speech recognition, (text-to) speech generation and voice control have been introduced the past decades and they are improving every day. However, can we enable people to create, access, utilize and share knowledge on the Web using only voice? Can we make it work for many small local languages? Can we make systems for people with limited computer-skills [8]?

This implies a number of interesting research challenges: (1) Can we create Web services that are accessible without hands? (2) Can the text-based Web and Semantic Web be augmented or even be replaced by a voice-based one? (3) How do we create, transfer, manage and query information based on speech? (4) Can we transfer Web knowledge to voice and vice versa? (5) How will we manage the voice-based knowledge on the Web? (6) Can the voice-based Web exist next to the text-based Web? (7) Should we create language- and context-dependent Web content for exchanging general knowledge? (8) Should we create language- and context-dependent Web content for exchanging local and specific knowledge? (9) How do we create automatic language and dialect classification?

3 Challenge 2: Have a good conversation with the Web

After enabling voice-based management of knowledge on the Web, we have to think about having a conversation with the Web or smart entities representing it. Keyword-based search strategies are too limited and semantic search strategies are fairly complex. Simple questions that involve a form of context, such as time, personal or professional situation, are still very hard to answer for current querying systems. Typical conversations between peers include: “*what is the best X for me*”, “*teach me how to Y*” and “*how do I feel better after Z*”. Current querying strategies involve searching for keywords - in this case X, Y and Z - and apply these in queries, ignoring the context.

In order for a system to have conversation with an end-user, it should be able to understand the context of that user. For example, a farmer in the Sahel is likely to have a different conversation than a stockbroker in Amsterdam. Furthermore, the system should be able to ask questions back or ask for feedback in case of uncertainty. For example:

User: “*Hi, I want to know what X is?*”. System: “*Okay, do you mean by X? X1, X2 or X3?*”, ...

User: “*Good morning, can you teach me to do Y?*”, System: “*Of course, How much time do you have available ?*”, ...

User: “*..., I feel really down, after Z?*”, System: “*Hmm, would you like to talk about Z?*”, ...

These kind of conversations should not be confused with Eliza-like chat bots, rather with expert systems combined with conversational knowledge. Next to that systems could contain a personality in order to have a “grown-up” and “social” conversation with the Web. Although many efforts, social speech is something that is both useful and currently unavailable. We believe that modeling *speech acts* is a first step in creating these type a systems. The original idea of speech acts is that conversations can be analysed on three levels: (i) the locutionary act: the actual generation and processing of speech, (ii) illocutionary act: the intended meaning and (iii) the perlocutionary act: its actual effect, such as persuading, convincing, inspiring, etc. [7]

Research challenges include for conversation systems include: (1) Can we create a voice based yellow pages services? e.g. “*connect me via telephone with an (human or digital) expert on matter X*”. (2) What kind of assistance can be performed only by dialogue? e.g. existing healthcare service can be consult via telephone, however, professionals can also detect whether it makes more sense to do a visual diagnosis. Can systems also perform vocal assistance and determine their own boundaries? (3) Will people trust these type services? Typical trust relations with human experts are based on reputation, education and previous experiences. Can we also do that with digital experts? (4) What are the limits these services? E.g. first thing the service could do is assess the type of task, e.g. classify (“*what is X?*”), design (“*advise me on Y*”), perform (“*pay Z*”) etc. Next it should ask for more information if possible? (5) Finally, would these systems perform better when having a form of personality formed around the personality of its users?

4 Discussion

Our ideas about voice and digital assistance are almost as old as the field of computer science. These are well documented by several visionaries and seen in science fictions. However, few examples exist that in fact fulfil parts of these visions. Remaining challenges include: (1): *I want to have a conversation with the Web*, (2) *I want to learn something from the Web, that is beneficial for me*, (3) *I want the Web to understand my feelings* and (4): *I want to make the world a better place, using the Web*. When tackling the above challenges, we could finally construct Hals, Datas or (positive) Marvin's being real vocal gateways to the knowledge on the Web.

Inspiration for above mentioned research origins from the Web Alliance for Regreening in Africa (W4RA) project and the VOICES [10] project. The W4RA project aims to bring the advantages of the Web to remote areas in Africa, making use of mobile telephony and a voice Web interface. One of the use cases in the VOICES project is about farmers in the Sahel that have two main sources for exchanging information: mobile telephony and community radio. Community radio is used to broadcast information. Old-fashioned mobile phones are used only for chatting. We believe African farmers and the other 4.5 billion internet-less people using voice-based interfaces will be able to benefit from the Web.

The voice and conversation challenges will bring new types of multi disciplinary research and technological spin-offs that will step far beyond the currently keyword, tagging, hyperlink-based and Semantic Web. These outcomes will turn the Web from a passive knowledge rich-environment accessible for few into an active supporting personalized environment that is accessible for all. The Web is World Wide. But precisely that is still the challenge.

References

1. The Economist, September 2009 Special report on Telecoms in emerging markets.
2. International Telecommunication Unit (ITU), 2011. Measuring the Information Society, Geneva Switzerland.
3. Mobile Web for Social Development Roadmap, W3C, 2009, <http://www.w3.org/TR/2009/NOTE-mw4d-roadmap-20091208/>
4. http://en.wikipedia.org/wiki/List_of_fictional_robots_and_androids
5. <http://www-03.ibm.com/innovation/us/watson/what-is-watson/index.html>
6. Berners-Lee, T. Fischetti, M. Weaving the Web, 1999.
7. Searle, J. R. (1969). Speech acts: An essay in the philosophy of language. Cambridge, UK: Cambridge University Press.
8. Lerer, A., Ward, M. Amarasinghe, S. : Evaluation of IVR Data Collection, UIs for Untrained Rural Users." ACMDev, London, 2010.
9. Akkermans, Hans and Gyan, Nana Baah and Bon, Anna and Tuyp, Wendelien and Grewal, Aman and Boyera, Stphane and Allen, Mary (2011) Is (Web) Science Ready for Empowerment? pp. 1-7. In: Proceedings of the ACM WebSci'11.
10. <http://www.w4ra.org/> and <http://www.mvoices.eu/>
11. Hendler, J., Shadbolt, N., Hall, W. Berners-Lee, T. and Weitzner, D.. 2008. Web science: an interdisciplinary approach to understanding the web. Commun. ACM 51, 7 (July 2008), 60-69.