# IMPROVING ACCESSIBILITY TO GOVERNMENTAL FORMS[1]

**Norbert Kuhn, Stefan Richter, Michael Schmidt, Andreas Truar**

Institut für Softwaresysteme in Wirtschaft, Umwelt und Verwaltung
Fachhochschule Trier, Umwelt-Campus Birkenfeld, Postfach 1380, D-55761 Birkenfeld, Germany,

{n.kuhn|s.richter|s.naumann|m.schmidt}@umwelt-campus.de

**Abstract:** Although many governmental institutions have provided their costumers with access to electronic government documents there is still a lack of accessibility for handicapped citizens. In this paper we present an approach to improve access to governmental forms for handicapped citizens, in particular for people with visual impairments, elderly people, illiterates or immigrants. We describe a system where we combine scanned images of paper-based forms containing textual information and a text-to-speech synthesis to an audio-visual document representation. We exploit standard document formats based on XML and web service technology to ensure independence from software and hardware platforms.

**Keywords:** Multimodal User Interfaces, Accessibility, Forms Reading for E-Government

## 1. Introduction

In recent years much effort has been spent in Human Computer Interfaces to improve access for handicapped persons to computer systems (Muller et al. 1997). To a major extent these activities are enforced by legislative constraints that exist in the US (e.g. the Americans with Disabilities Act (United States of America, 1990)) as well as in the European Union (European Commission, 2000), and in its member countries, like in Germany (Bundesrepublik Deutschland, 2006, Bundesrepublik Deutschland, 2002). However, a large amount of these realizations allow the user only to download particular forms, to print them, and

[1] This work is supported by the German Ministry of Research and Technology under grant FKZ 1771X07

send it back to the governmental institution after some information has been inserted. Up to now, in many countries the use of printed documents remains the dominant means for the communication between governmental institutions and citizens.

While for the web based information systems accessibility aspects are often considered in E-Government platforms, for the procedure of forms completion in most cases it is necessary to process printed documents, yielding a point of media disruption which is difficult to handle for many users with particular handicaps.

In this paper we present an approach to build interfaces to governmental forms. Our main target is to develop tools to be able to cope with printed documents. However, some of the ideas can also be applied to electronic documents. Therefore, we exploit different computer science technologies e.g., from the fields of document analysis, language processing, and distributed systems to develop our solution.

The work presented here is embedded in the FABEGG project (Framework for modeling and acceleration of E-Government applications processes) and the GUIDO project (Generating user-specific interactive documents) which are both granted by the German Ministry of Research and Technology. In these projects we are concerned with the human computer interfaces to E-Government information systems. We deliver general guidelines how to build such interfaces, and develop generic concepts and modules that should be available for different platforms. We base our document representation on XML structures and implement communication by using web services. This guarantees independence from software and hardware platforms.

Our prototypical system used to demonstrate and to evaluate our work is called EG-VIP. In most cases when we speak about documents within this paper we have in mind governmental forms, which are used to provide and to maintain information that is necessary to run a governmental process. Such forms may be partially filled already, either by the authority itself or by other persons who interacted with the document before. In particular, when information from a citizen is required the authority usually sends a printed form which has to be completed and returned to the governmental unit where the next step in a process is executed. Thus, a form is often an instance of some form template to which individual data identifying the recipient and the corresponding process is added. This is e.g. the case for mail questionnaires in legal processes, real state affaires or tax proceedings.

## 2. System Architecture

Our main target is to develop interfaces to E-Government systems which are able to handle governmental forms. Moreover, we consider for this task users suffering from reading disabilities. On the one hand, these are people with visual impairments, ranging from colour blindness, over low vision to blindness. Another group of users we have in mind are those who suffer from dyslexia. All these users imply quite different requirements to the design of the interface and the interaction with an information system. In order to be able to economically develop adequate interfaces the system should adapt almost automatically to the actual user's needs. Figure 1 gives an overview of the system architecture we have chosen for that.

The EG-VIP system consists of different components. The core component is the GUIDO server in the lower left corner of the figure. It is basically a document database where all the documents of a governmental authority may be stored. A database entry for a document consists of different data. This comprises an image of the document which can be used to print it or display it on a screen. Furthermore, we expect to have a semantic description of the document, including information about the fields of a form, their type, or possible relations between them, which ones will be filled already with information and which ones have to be completed by the addressee. Also a link to the workflow management procedure in the authority should be present, i.e. a pointer to the description of the workflow where the document is used for. We call the universe of information associated with a governmental document or a form the Generic Document Structure. The document repository with all the generic document structures is built by the governmental agencies or associated business companies, e.g. the utility company or the water works.

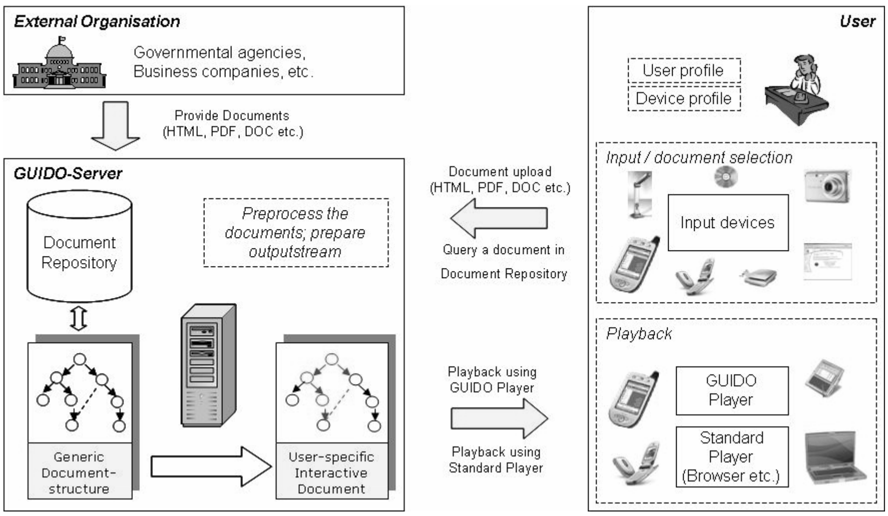


**Figure 1. Architecture of the EG-VIP system**

When a user interacts with the server site, the EG-VIP system derives an appropriate representation of the governmental documents for the user. We call this representation the User-specific Interactive Document. To compute the latter one from the generic document structure the system needs information about the user's sensory and physical capabilities and about his preferences. This will determine whether the server will run OCR or not, includes audio information, or computes transformations of the image which is actually processed. The choice will be based on the user profile, the actual bandwidth available for the client-server communication and on the features of the user device, which are given by the device profile. If the client computer has no speech synthesis program available the audio information has to be computed by the server. Bandwidth parameters can then determine the compression rates of the audio and the graphical information.

## 3. Constructing the User Interface

The right-hand side of figure 1 shows different devices with which a user may interact with an authority portal. This includes a desktop computer, or a smart-phone. Other devices may also be used, but the figure includes only devices which are able to start interaction with the server as well as to receive the corresponding answer. However, this assumption is not necessary and could easily be discarded.

At the beginning of the communication the user sends an image of the form he wants to work on to the server. He may take a picture either with the device itself (e.g. the camera within the mobile phone) or with a separate device, for example a digital camera or a scanner. Together with the image, the device and the user profile are sent. The device profile comprises information about the features of the device sending the request, e.g. the size of the display (height and width), or information about the software which is installed, for example whether an OCR program or a speech synthesis tool are available on the device. In addition, the service considers which software is used to present the result to the user. For example, this may be an EG-VIP specific software component allowing more interactions for the user compared to the situation when a general tool like a browser is used.

The user profile defines personal preferences for the visual or the audio result of the request. This may concern for example the color combination of the document which will be displayed by the device (e.g. black-white, black-cyan, cyan-yellow, or none). For the audio result the compression method or the sampling rate can be specified. Both, user profile and the device profile determine the result which is expected from the server. They may vary for the same user according to the situation or the location in which he currently is. For example, he may have installed an OCR-software on his desktop computer but not on his mobile device, or he prefers to receive only the audio output when he is traveling around, while at home he wants to receive a black-green document which he can enlarge and read himself. Both kinds of profiles are described in an XML-file which is transferred to the server as part of the service request.

When the server receives such a message it first has to analyze the document, i.e. the server tries to find an appropriate generic document structure within the document repository. For this task we currently use the result of an OCR process giving us specific keywords which are used to retrieve a document description. Other approaches, for example barcode recognition could also be useful for this purpose.

After the document has been successfully identified its generic document description is read which contains information how to process the image further, e.g. it describes which information can be expected to be already available in certain fields and hence, has to be extracted by the OCR process and which information should be requested from the user. The server packages all the relevant information into the answer it returns to the user client. In order to achieve platform independence and to be able to incorporate into our scenario all the devices shown in figure 1 we encode the answer into an XML file format. Figure 2 below shows an excerpt of such a file.

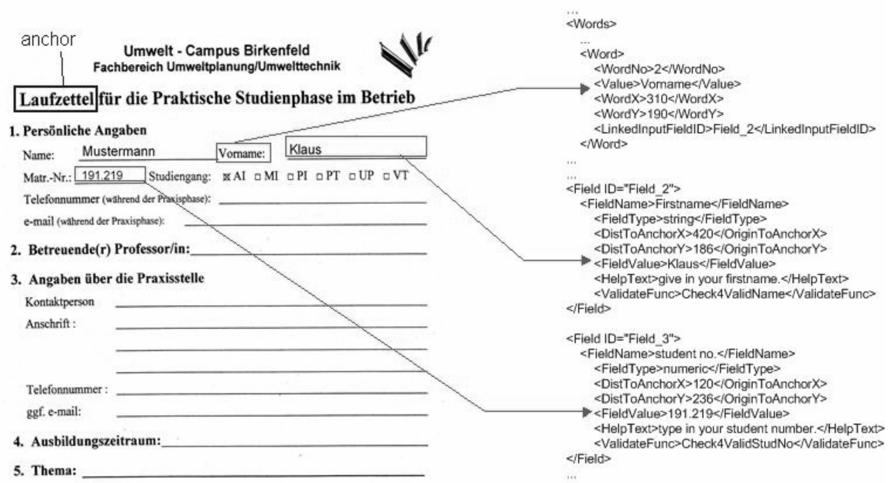


**Figure 2. A Governmental form and its representation in the user-specific document**

On the left-hand side of figure 2 you may find a form from our local university administration. It contains already the name, surname, and the registration number of a student. This information is extracted automatically from the document using the information which fields have to be processed in this and where they are located. This data is encoded in the XML-document as shown on the right-hand side of figure 2.

The structure of the result is almost the same in any use-cases. However, there are different ways to use this structure. In the following we want to discuss two scenarios.

### 3.1 Example 1: Desktop Client

Figure 3 shows a high level configuration of a user client system running on a desktop computer. It consists of a camera unit -shown on the right-hand side - that allows to capture printed documents and the display with speakers built-in. The camera unit enables easy document handling and fast document capture and moreover, is able to treat documents of larger size. Alternatively a scanner could be used. Such a user terminal could be placed for example in a major administrative department. In this case the speakers would be headphones due to privacy considerations.

**Figure 3. Desktop client with digital camera**

The system can further vary according to the software which is installed. For example, we may have OCR-software, as well as speech-synthesis and speech recognition software locally available resulting in a highly autonomous system. This setting corresponds widely to the features of our first prototype which is referred to for example in (Kuhn et al. 2007a) and (Kuhn et al. 2007b).

In this environment it is possible that after the document has been processed by the local OCR software it can be read by using the speech synthesis software. The software component we use for speech synthesis also allows text reading in different velocities. While reading the text the system highlights the word which is actually spoken by drawing a (maybe coloured) rectangle around it. This enforces auditive understanding. With a pointing unit (e.g. a mouse or the finger when a touch screen is available) the user can start and stop reading at any arbitrary position. As the system is also able to process several common electronic formats like PDF or HTML, it is also an appropriate means for providing access to general electronic information systems.

With this configuration communication between the client and the server is only necessary when semantic information about the document is asked for. So the server can provide the client with type information about the fields on a form or with information related to a process where the form belongs to. In turn, this information can then be used to verify data from the OCR software which have not yet been recognized with sufficient confidence or to check user input for validity. Another important type of information which the server can deliver are help texts to guide the user through the process of completing the form and explaining to him which information is required in a concrete situation. The help texts are also present in the XML document of figure 2.

If we have a client at our disposal where less software features are installed the server is urged to provide the lacking information. Usually we encounter this situation when we use a client with less computational power, like a mobile phone or a smart-phone.

### 3.2 Example 2: Mobile Client

In 2006, Kurzweil Technologies Inc., an American manufacturer of assistive products and the National Foundation of the Blind presented a mobile phone equipped with OCR software which allows reading texts with that phone which are captured by the integrated camera (Kurzweil Technologies Inc., National Foundation of the Blind 2008). Recently, in cooperation with Nokia similar software has been developed and can be bought for a Nokia mobile phone, which is much lighter and more powerful than the original knfb reader.

This product yields a substantial improvement for the autonomy of humans with low vision when they are en-route away from the equipment in their home office. However, the solution is currently bound to single hard- and software system. The approach we present here overcomes this restriction and can in principle be run on any device which has available an internet browser. Nowadays this is true for the vast majority of the mobile phones which are being sold.

Figure 4 shows the representation of a German governmental form on a smart-phone emulator from the MS Visual Studio .NET development framework. We used this framework to implement the mobile version of our EG-VIP system. Many of the interactions between user and interface known from the desktop client are also available on this platform now.


**Figure 4: Form representation on a mobile client.**

As you can see by the box around the text, it is possible to step through a document word by word and get the word which actually has the focus visually marked while it is read. It is also possible to choose among different color combinations or to enlarge the image. Furthermore, the service can deliver the audio information which represents the spoken text of the document. For this the user may choose among different modes, i.e separate the audio information into different streams for each word, or for each paragraph or receiving a single stream.

Unlike the Nokia/Kurzweil approach we execute most of the computation on the server. To do so, we have encapsulated our formerly local application into a webservice which can interact with the mobile client over a network (e.g. the internet) answering HTTP based requests. One advantage of this solution is that on the server we can apply more powerful software tools, e.g. for the OCR and the speech synthesis. Thus, we expect much better results concerning the text recognition rate and the quality of speech.

The computational task of the client is restricted to the interpretation of the user-specific interactive document (cf. figure 1). As mentioned before this is an XML document where the client can find any data which is necessary to display a document and which the client can not (or does not want to) produce itself. In an extreme case the XML document can degenerate to an HTML document, containing for example only the document in a certain color combination or represented as an audio file which is played by a browser.

## 4. Conclusion and Outlook

The EG-VIP system provides humans with reading disabilities with much better access to information systems. In particular, access to printed documents, which are still important in business processes or in governmental applications, is substantially improved. Moreover, our solution can also be used in every day situations, e.g. when a visual impaired person stands at a bus stop and needs to read the schedule there.

We have further discussed our ideas to automatically adapt the output of a centralized information portal of a governmental authority to the special need of a citizen looking for information. The core of our system is the document repository where the different documents are maintained and the user specific interactive document is generated. The implementation of both a desktop client and client for smart-phones gives us strong support for the viability of our approach.

In our projects different partners are involved, among them a governmental authority and a German manufacturer of assistive technologies. Together with them we will integrate our approach for dynamically constructing the document interface and evaluate it in field tests together with citizens and handicapped users.

Currently, our main focus lies on the group of users with visual disabilities. This group of users will increase in number at least in many European countries due to the demographic development. A further possible target group are people with difficulties in reading comprehension, e.g. immigrants or dyslexic people. Studies from the OECD have stated that all over the world there are 800 million people who are not able to read or write. Even in the Western European countries the percentage of the inhabitants showing functional reading disabilities is estimated to be more than 15 percent (International Adult Literacy Service, 1998). Providing them with the automatic reading of documents and including in addition a speech recognition facility into the system will allow them to access to governmental processes more easily.

## References

Bundesrepublik Deutschland, 2006. Allgemeines Gleichbehandlungsgesetz der Bundesrepublik Deutschland vom 29.06.2006, http://www.gesetze-im-internet.de/agg/

Bundesrepublik Deutschland, 2002. Verordnung zur Schaffung barrierefreier Informationstechnik nach dem Behindertengleichstellungsgesetz (Barrierefreie Informationstechnik-Verordnung - BITV) vom 17. July 2002

Edwards, A. D., 1988. The design of auditory interfaces for visually disabled users. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Washington, D.C., United States, May 15 - 19, 1988). J. J. O'Hare, Ed. CHI '88. ACM Press, New York, NY, 83-88.

European Commission, 2000. Gleichbehandlung Behinderter in Beruf und Bildung: Richtlinie des Rates 2000/78/EG vom 27. November 2000, ABl. L 303 vom 2. Dezember 2000.

International Adult Literacy Service, 1998, Report for the Organization for Economic Co-operation and Development (OECD), http://www.statcan.ca/english/Dli/Data/Ftp/ials.htm

Kuhn, Norbert; Richter, Stefan; Naumann, Stefan: Improving Access to EGovernment Processes. In: Khosrow-Pour, Mehdi (ed.): ManagingWorldwide Operations and Communications with Information Technology. Proceedings of the 2007 Information Resources Management Association International Conference (IRMA 2007) Vancouver (British Columbia), Canada (2007a). IGI Global, Hershey / New York 2007, pp. 1205-1206

Kuhn, Norbert; Richter, Stefan; Naumann, Stefan: Improving Accessibility to Business Processes for Disabled People by Document Tagging (Poster Presentation). Proceedings of the Ninth International Conference on Enterprise Information Systems (ICEIS 2007). Funchal (Madeira), Portugal (2007b), pp. 286-289

Kurzweil Technologies Inc., National Foundation of the Blind: Reader software for a mobile phone, http://www.knfbreader.com/index.php (15.02.2008)

Muller, M. J., Wharton, C., McIver, W. J., and Laux, L. 1997. Toward an HCI research and practice agenda based on human needs and social responsibility. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Atlanta, Georgia, United States, March 22 - 27, 1997). S. Pemberton, Ed. CHI '97. ACM Press, New York, NY, 155-161.

Richter, S., 2003: Design and Implementation of a communication module for blind and visually impaired humans. Diploma thesis, Birkenfeld.

Spiegel Online: Handy für Blinde liest gedruckte Texte vor. http://www.spiegel.de/netzwelt/mobil/0,1518,532014,00.html (15.2.2008)

United States of America, 1990. Public Law 101-336, 1990. Text of the Americans with Disabilities Act, Public Law 336 of the 101st Congress, enacted July 26, 1990.

W3C, 2006: Guidelines and resources from the World Wide Web Consortium (W3C). www.w3.org/WAI/