MobileSage — A Prototype Based Case Study for Delivering Context-Aware, Personalized, On-Demand Help Content

Till Halbach & Trenton Schulz
Norwegian Computing Center
Oslo, Norway
Email: {till.halbach,trenton.schulz}@nr.no

***Abstract*—We present the system design decisions for the MobileSage Prototype, a service for the on-demand delivery of multimodal help content to anyone in general and seniors in particular. Findings from user-centered research formed the system requirements, design considerations and decisions. The design of the system also includes availability, relevance, accessibility, conciseness, and comprehensiveness of multimodal content. The prototype has been evaluated in user trials with encouraging results, showing the participants' high appreciation of the system.**

Keywords*—*Mobile; smartphone; application; assistance; guidance; help on demand; personalization; adaptive; accessible; usable; multimodal; context; location aware; Ambient Assisted Living.

I. Introduction

Our everyday lives are becoming ever more technological with an increasing number of machines and devices surrounding us. This leads to a continuously rising degree of complexity in everyday life. This can be a challenge for many – particularly elderly persons. This is why many senior citizens meet these solutions based on Information and Communication Technology (ICT), such as ticket machines and web services, with anxiety.

At the same time, modern elderly – here defined as people aged 65 and older – live longer, are healthier, more active, mobile, independent, and more demanding customers [1]. There are approximately 87 million elderly in Europe [2]. They are increasingly looking for useful, user friendly, and personalized ICT services that add value to their active and mobile life; they also desire services that can help them to stay active despite various impairments. MobileSage provides a timely approach and solution that is detailed below.

The article is organized as follows: First, we present the system architecture of MobileSage and its main components. Next, we lay out the requirements for these components and their consequences for system design and content production. We then present results from the first round of user evaluations in Norway. Finally, we conclude with next steps in the project and other lessons learned. This work's main contributions cover the thorough discussion of the needs of various user groups, the derivation of constraints for the design of the system and the production of media content, as well as results and recommendations from user trials.

II. MobileSage Overview

MobileSage stands for *S*ituated *A*daptive *G*uidance for the Mobile *E*lderly. Its aim is to give the modern elderly a smart agent that provides relevant, accessible, usable, and multimodal assistance for carrying out and solving everyday tasks and problems in the self-serve society, whenever and wherever they occur [3, 4]. Related research and solutions are discussed in [5].

A. MobileSage Components

MobileSage consists of two components: the Help-on-Demand (HoD) mobile application and the Content Management Service (CMS). Figure 1 shows the overall architecture. The system is developed in three major iterations, and the findings here are from the first iteration.

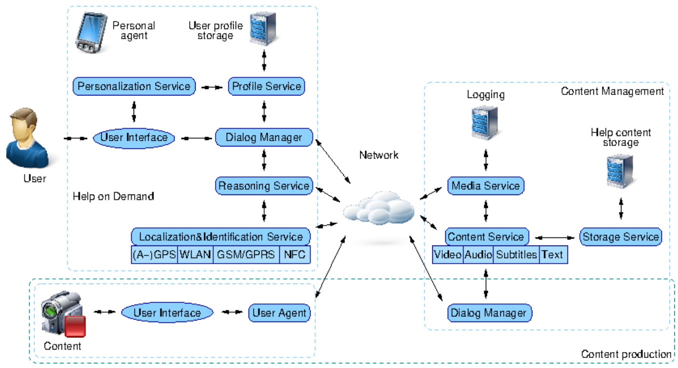

Fig. 1. System architecture and major building blocks for HoD (left) and CMS (right)

1) Help-on-Demand Service: The HoD application is the personal agent, a thick-client application running on a smartphone. It is built up in a service-oriented manner, see Figure 1. The user interacts with the Dialog Manager through the User Interface. The Dialog Manager utilizes information from the Profile Service taking care of the user profile. The user profile stores personal preferences and usage patterns. User behavior and User Interface events are logged and analyzed by the Personalization Service, upon which the user profile is readjusted.

The Dialog Manager is in contact with the Reasoning Service to help determine the user's context. Reasoning makes use of network services such as Media Service, Search Service, and the Content Service. The Reasoning Service gets help from the Localization Service, which can determine the user's location based on technologies like A-GPS, WLAN, GSM/GPRS, NFC, and triangulation methods.

The HoD Service requests the content from the CMS upon initiation of the user.

2) Content Management Service: The CMS is a cloud service running on a web server. Content producers interact with the service's Dialog Manager, which in turn controls the User Interface on a User Agent like a web browser. The logic for handling the multimodal content lies in the Content Manager, which has a modular design to be able to add additional modalities in a simple way. The prototype supports the modules Video (with or without captions), Image, Audio, Text, and Formatted Text (basically simplified HTML). The content is stored by the Content Service. It is also possible to refer to content located elsewhere (e.g., from other video services).

There is no limitation in the kind of content that can be facilitated. This includes manuals, usage instructions, descriptions of travel routes, and geographical points of interest. We anticipate that vendors or service providers generate most of the content. For instance, a particular vendor might provide manuals for their ticket machines, or the railway operator that runs these machines might do so. Even a municipality might be interested in producing such help content as a special service for their citizens and visitors. Other interested parties are expected to add content to the CMS to fill in the gaps left behind by vendors or service providers, as they are likely to have a direct interest in helping someone using HoD. Finally, there is nothing that prevents users of the HoD from producing and making help content available themselves.

III. User and System Requirements

The following sections address the formulation of the requirements and constraints for the system design. The primary MobileSage users are those using the HoD service. Secondary users are content producers, and tertiary are vendors or service providers.

A. Requirements for Primary Users

The derivation of the requirements of primary users is split into the gathering of the users' expectations towards the system (user needs analysis), and the collection of user requirements. The system requirements were derived from the latter.

1) User Needs Analysis: Focus group work was conducted in the three countries Norway, Romania, and Spain to find the needs of primary users [6]. The focus groups had 39 participants and represented a broad range of parameters, including age (48 to 96), gender (24 female vs. 15 male), disabilities (sensory and cognitive impairments), nationality (4 foreigners), and ICT experience and usage. Two scenarios were presented to the participants: an individual with reduced vision traveling in a foreign country where he was not proficient in the language, and an elderly lady at home trying to understand how to use an electric household appliance.

The focus groups' results show that "modern elderly persons" is a very heterogeneous group with a wide range of – sometimes contradictory – needs and wishes. This applies also to the users' familiarity with ICT in general and mobile technology, which ranges from none to professional users. However, it was possible to identify themes of what the solution should have [6]. The solution should *a*) lead to higher independence of elderly people according for the help-for-self-help principle, *b*) increase a person's mobility, and be usable for transportation and travel, including holidays and visits, *c*) be applicable in the home environment and throughout daily living, *d*) provide relevant, useful, context- and location-sensitive and multimodal assistance in an on-demand manner, *e*) be accessible, user-friendly, designed for all, possible to personalize or customize, adaptive, social, and *f*) honor privacy, security, and trust matters.

2) User Requirements: The results from the user needs analysis were collected and formulated as user requirements [7]. The roughly 50 requirements mirror the expectations of primary users regarding HoD, but were extended to be valid for the CMS as well. The user requirements served as input to the process of formulating the first draft of the system requirements for the service's two components.

3) System Requirements: The requirements for the Help on Demand and Content Management services were derived from the user requirements.

The requirements specification for HoD has over 60 requirements [8], while the CMS specification contains only 40 [9]. Both address topics such as system functionality, user interface, and input and output matters. Also included are sections on the technology choice and mockup examples regarding the services' user interfaces.

B. Requirements For Secondary and Tertiary Users

MobileSage's focus is on primary users. Secondary and tertiary users have been accounted for by formulating a set of requirements representing the needs of the transport company participating in the project. These are as follows: *a*) It should be possible to identify one or several points of interests with a unique ID. *b*) There could be multiple help topics per ID. *c*) One topic could be presented in multiple languages. *d*) The service should support content hosted elsewhere ("upload once, available everywhere"). *e*) It should be possible to edit help content in order to add locations, languages, and modalities.

IV. System design

For the HoD service, a user profile lays the ground for personalization and adaption of the service. It contains the user's settings and preferences, such as font size, emergency number, accepted media types, and additional languages. Also other parameters are stored there, including usage log. This log is the basis for system adaptation. Screenshots of the HoD are shown in Figure 2.

Both primary and tertiary users have requested that it should be possible to associate content with specific locations or points of interest. However, it should also be possible to link certain content to several locations (e.g., "how to buy a ticket" is valid for any ticket machine in the Oslo area). Moreover,

 ISBN: 978-1-61208-306-3

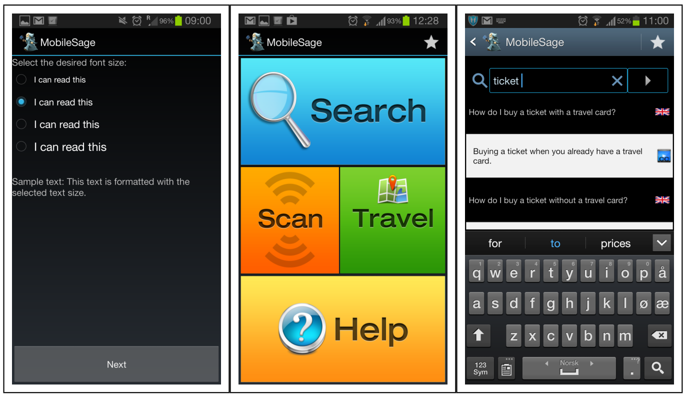

Fig. 2. Screenshots of the Help-on-Demand application: set up (left), home screen (middle), and search (right)

there are situations where several pieces of content are relevant at a single location (e.g., how to validate a ticket, arrival time of the next bus, or choosing the correct platform for departure).

These issues have been solved by the Content Item, see Figure 3. This uniquely identifiable item is a logical unit to gather content that is related to each other. Multiple locations in terms of latitude, longitude, and altitude can be linked to a single Content Item, and so can Records, each representing a particular topic. The topic itself is described by a Record's title together with a language identifier. Language translations of a topic become a new Record. To avoid topics being mixed in the result listing, the results are ordered according to language first, and alphabetically by topic. The user needs analysis recommended further to split content into several Steps or segments, and to promote segmented content, something the presented data model is capable of by combining multiple Steps into a Record. A Step has the same language as the Record it belongs to and has one of the Media Types: text, formatted text (HTML), audio, an image or animation, video, or a video with captions. Further elements are a brief Summary and an URI/URL pointing to the media itself. The URI can point to a server that is part of the CMS, but it may also point to external resources (e.g., a video on YouTube). For such external resources, the CMS effectively functions as a service holding metadata on indexed resources. This model supports multiple media types not only for the same Step but also mixing of media types per Record (i.e., several Steps) depending on which type suits a particular step best. For instance, video might be best suited to illustrate a movement, while often a still image is beneficial for highlighting a specific region of a visual.

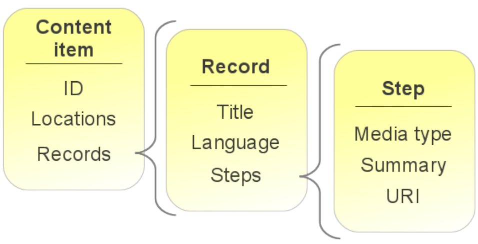

Fig. 3. Data model of the content

MobileSage is about just-in-time guidance and *on-demand* assistance. Based on suggestions from the primary-user studies, it was deemed too intrusive to let the mobile application initiate requests for help based on the location of the phone at points of interest, nearby radio fields, etc. Thus, the user indicates a wish for assistance either by scanning a QR code orNFC tag, or by sending a text phrase to the CMS. In the former case, the code or tag carries the ID of a particular Content Item, which is read by the mobile app and sent to the CMS, resulting in a list of all topics associated with that ID. Regarding the search phrase, topics are viewed as relevant regardless of the ID, accounting for both Record titles and Step summaries.

One of the challenges of MobileSage is to find *relevant* content and not to confuse the user with extraneous information [9], which helps individuals with orientation and problem solving challenges. The key to this problem is determining the user's context, in terms of (most importantly) location, time and date, user habits, and other aspects. Nearby objects are considered relevant in the CMS by calculating a proximity radius around the user's current position; only Content Items with a location within this circle are returned as results to the phone. The exact radius of the circle was based on heuristics and set to roughly 40 m.

Records are sent in "pages" to the phone, meaning that HoD tells the server how many results per request to return. This is done first of all for practical reasons, i.e., bandwidth limitation, and second because the user is likely to be interested only in the most relevant results, which are presented first. The client on the phone keeps track of the number of transmitted records and is hence able to request a particular page with results, say, page 3 with the records 21 through 30 in case of 10 allowed records per page. If the user scrolls down while being at the bottom of the results list, the client fetches more results if available.

For simplicity, any media content is offered to the HoD as a file for download through HTTP. While this works great with text-based content, the performance in terms of responsiveness of the playback on the media player is suboptimal when connected over a channel with very limited capacity, as discussed in Section VI-B because media downloads in most clients have to finish before the media is rendered on an output device. Clients that support (true) media streaming and pseudo streaming methods like HTTP Live Streaming will start rendering the output as soon as sufficient data become available. These methods require the proper setup of a streaming media server and are planned for the next iteration.

V. CONTENT PRODUCTION

This section considers the production of content for MobileSage in particular and educational and instructive content in general.

The content found should be relevant, concise, and comprehensive. However, as recent research surveys show [10–13], it is extremely difficult to develop methods which can check exactly that in a satisfactory manner. MobileSage offers a manual approach in its CMS [9]. As mentioned before, the splitting of longer content into shorter steps is encouraged. The content producer now provides the content abstraction on two levels: A summary of the step itself, and a title wrapping-up of the entire record (see Figure 4). The content producer must tag the content with the proper descriptors for language, and its location, if applicable.

Currently, the content must be uploaded in a format accepted by the Android OS. This applies to both the format of the content tracks, be it video, audio, or captions, and the format of the embedding media container. In future iterations, the CMS is planned to accept any format with transcoding into a proper format supported by the OS. This targets locally stored and remote content likewise.

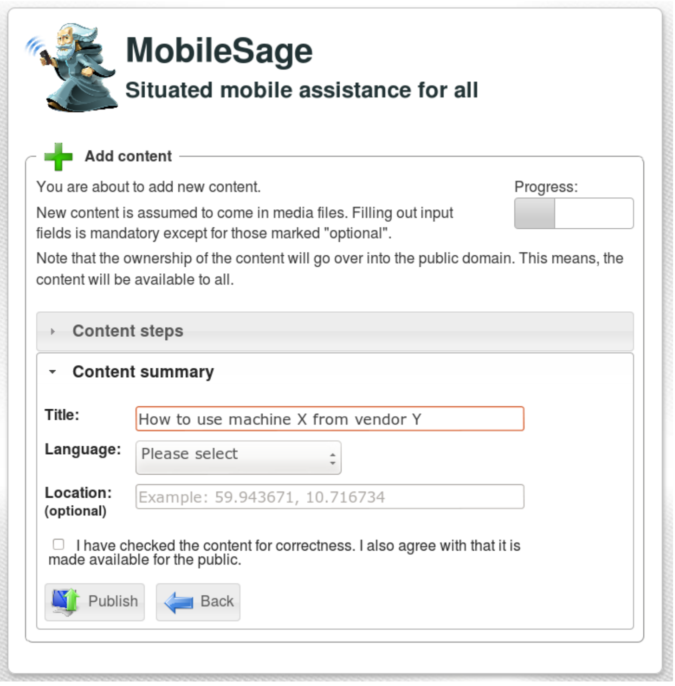

Fig. 4. Screenshot from the uploading of new content to the CMS.

While the video material could be presented at any resolution, we chose to encode video with a resolution of 480×360 pixels and a bitrate of roughly 200 Kbps @12 fps in H.264 Baseline Profile Level 2.2 format, and embedded in an MP4 transport container. The audio tracks (both stand-alone and as part of a video) had a rate of roughly 48 Kbps @22 KHz mono and were encoded in AAC format. A video containing one visual and one audio track thus had a bitrate of roughly 260 Kbps, which includes overhead data used for track muxing and the container format. The length of the content varied from approximately 10 s to 4 min, but most of the content was between 30 s and 45 s.

We used open-captioning, where a voice's transcript are printed as always visible titles on the screen, instead of composing a separate captions track. This avoided the extra work of producing a captions file and ensured that the video player always worked properly, without any user action. We chose a “slab serif” font that was originally designed for fax machines, with a size of at least 36 points. One disadvantage to this approach is that more space is taken up on the content server to store each captioned video, but as the videos were short and at a low resolution, we believe this is a minor issue.

The content is currently provided in a single quality in terms of resolution/sampling frequency and bitrate as mentioned above. Both have implications for the bandwidth necessary to transmit a given content file to the client's media player: the larger a picture (in pixels), and the higher the audio sampling frequency (in Hz), the more bandwidth will be necessary. Likewise, the higher the encoder bitrate (in bps), the more bandwidth is required. Channel capacity of the cellular link, however, is a limited resource for physical reasons. It takes time to transmit a particular amount of data over a channel, which also has an impact on the user experience. The service is thus required to respond upon user interaction within a reasonably short time span [7]. Currently, this requirement is not reflected in the system, but it is planned to honor it by measuring the duration of packet downloads at the client side (which may vary over time according to the signal strength and coverage) and include this information in the server requests, together with information about the phone's screen size. In the next development iteration, content will be provided in several resolutions or sampling frequencies, and several bitrates. Media searches at the server side can then utilize this information about the channel conditions and limit the results to media qualities that meet the bandwidth constraints. For example, a phone with a 480×800-pixel screen is connected to the network over a GSM (2G), and the bandwidth averaged over 20 s is measured to be 100 Kbps. Based on the different resolutions and bandwidth, the server decides that a 480×360-pixel will still render acceptable to good quality. Yet, the 480×360-pixel video comes in two different bitrates; one encoded at a rate of 260 Kbps, and one encoded with 130 Kbps (assuming a constant encoding bitrate). The latter is closest to (but still above) the estimated channel capacity and will be sent to the phone to minimize the service's responsiveness, together with a notification about poor channel conditions.

VI. User Evaluations

User trials were conducted in Norway, Spain, and Romania in December 2012 to evaluate the service after the first development cycle.

A. Setup

The evaluation in Norway consisted of a travel situation at a subway station in Oslo, where participants used the prototype to find help with getting to a subway station, finding a ticket machine, buying and validating a ticket, and choosing the correct platform. We created content for all of these scenarios, with a minimum of audio and video for each. All but one of the scenarios had captioned video, and some had a textual modality in addition. Each of audio, text, and video was done both in Norwegian and in English to allow users to choose an additional content language. We then created several NFC tags for each of the scenarios and used it as a way of getting the information. We also wanted to use QR-codes, but we ultimately dropped it as it was to unreliable to scan with the current version of the mobile application. We tested on two smartphones with an Android OS 4.1 and screen sizes of 480×800 and 720×1184 pixels without any discernible difference in the results.

 ISBN: 978-1-61208-306-3

Eight participants were recruited for the evaluation. They were between 65 and 76 years old, and four of them had no experience with smartphones; however, they were somewhat experienced with computers. A few of them knew the location to a certain extend. A session consisted of a short introduction to the MobileSage idea, followed by a brief interview concerning their experience with mobile phones. Then, we demonstrated the application. After this, we began the tasks. One evaluator would take notes, while the other would guide the participant to make sure a task wasn't forgotten. After completing the tasks, there was a short follow-up interview about the service and the participant's experience about it.

B. Results

In the first task, the participants had to create a profile that matched their preferences for text size, language, and types of media they wanted to receive help in. The users understood the concept of several content languages, and the majority (75%) added English to their profile. The user-specific media types ranged from a single one to the entire range as detailed in Section IV, where captioned video was chosen most often. The majority (90%) of test persons checked 4–5 media types including audio, even though some participants said they would avoid wearing earbuds or headphones. Text was never requested. Choosing video and captioned video was inconsistent, hinting on a potential misunderstanding of the user concerning the meaning of "captioning". It is hence recommended to improve the description of media types or show brief examples of them. All participants but one expressed that making the profile was sufficiently easy.

The second task concerned navigation, where the participants had to get from their current location to the nearest subway station. All were able to enter the information needed, but the phone's positioning worked unreliable, sometimes placing the participant a block further south and/or facing the wrong direction. This issue sorted itself out when walking to the location.

The next task dealt with getting help at the ticket machine. Two participants were not able to finish this task due to a technical issue that caused no results to be returned from the CMS, which was corrected subsequently. All others succeeded with using NFC tags or by manually searching for information about where the ticket machine was, how to purchase a ticket, how to validate the ticket, and which platform they had to go to. Though only one was familiar with the technology, two had heard about it, and the rest were unaware of what it was, all really liked the technology and experienced it as easy to use.

A problem encountered was the effect the environment had on the signal strength in the phones. While above ground, it was possible to get video and audio without any issues, and the selected item would show up almost instantly. Yet, underground in the subway station, it became very troublesome for the phone to contact the content server. The main reason for this is that the only connections that are currently available in this particular station are so-called Edge (2G) connections. They are much slower compared to a 3G connection and also very latent. This was no big issue when retrieving, say, the results list. Participants had to wait a long time, though, if they wanted to watch a video. The audio fared a little better, but downloading would not always complete. Sometimes, the application on the phone would simply give up and it would be necessary to download the audio or video from the beginning. Most participants noted that it took a while to get the information in this case. With the continuing widespread of 2G connections in many countries, it is recommended to produce at least one version of low-resolution low-bitrate content, and to use techniques that increase the responsiveness of media players, such as media streaming, as discussed in Section IV.

No users complained about the size, resolution, quality, frame rate, or length of the video. Some participants noted that the font used for the captions indeed was sufficiently large and easy to read. There was only one instance where people commented on unclear information, where a video showed an unreadable display on a ticket machine.

All the participants felt that a help-on-demand system was something that would be useful for them. One of the participants even claimed that she was scared of using the ticket machine and always went to a store to buy her tickets, but would now she would use the ticket machine since she felt confident that she could use it based on the provided instructions. Concerning potential improvements, the most popular suggestions were a shorter response time for videos (when in the subway station) and dynamic information, such as time schedules. Those familiar with mobile applications suggested to include MobileSage's functionality in the public transport provider's current smartphone application.

VII. CONCLUSION AND OUTLOOK

We presented a first version of the MobileSage prototype, a service for delivery of context-aware, personalized help content in an on-demand manner. MobileSage incorporates the needs of primary, secondary, and tertiary users and has been been evaluated by primary users in user tests.

The aspects of content provision include multimodality and internationalization to take care of user personalization, multi-resolution and multi-rate for device adaptivity, and location-aware media searches for relevance. It has been shown that the system can index both internal and external media databases.

The idea of MobileSage was well received, as the participants in all focus groups and user tests viewed the service as useful. Yet, they also pointed out areas for future improvement: The concept of media types is not fully understood, in particular "captioning"/"subtitles" remains an unclear term to most users. Positioning could be improved, for instance by combining several techniques like A-GPS and WiFi triangulation. Also, users requested more information like transport schedules to be integrated into the service. The application needs further improvement in presenting results. The challenge is to find content relevant for the given context and reduce the need to search for the correct result. For example, only a single language and the "most rich" media type could be presented for a particular topic. Finally, the responsiveness in situations with limited bandwidth could be improved by using more advanced protocols and device and channel adaptivity.

Content producers need guidelines on how to produce content. While it is empowering that anyone can contribute content, it is important that the offered content is usable. Guidelines should include information about the length of help sections, phrasing of information, and how multimedia content should be organized. The results from the user trials can help with preparing these guidelines.

MobileSage will proceed its work with two more development iterations, combined by user evaluations, before the final version is going to be released early in 2014. In the next user trials, we are also planning to give special attention to how QR codes are accepted as compared to NFC.

ACKNOWLEDGMENTS

This work is partly funded by the European Commission through the AAL Joint Programme, the Norwegian Research Council, and national bodies in Spain and Romania. The authors would like to thank the consortium members for their valuable contributions and all individuals involved in the user studies for their feedback.

REFERENCES

[1] D. Metz and M. Underwood, *Older, Richer, Fitter: Identifying the Consumer Needs of Britain's Ageing Population*. Age Concern England, 2005.

[2] European Commission, "Digital Agenda: Commission proposes rules to make government websites accessible for all," retrieved 2013-08-20. [Online]. Available: http://europa.eu/rapid/press-release_IP-12-1305_en.htm

[3] V. Sánchez, "The MobileSage Project," retrieved 2013-08-20. [Online]. Available: http://mobilesage.eu

[4] I. Solheim and T. H. Røssvoll, "MobileSage Project factsheet," retrieved 2013-08-20. [Online]. Available: http://nr.no

[5] T. H. Røssvoll, "The European MobileSage Project – – Situated adaptive guidance for the mobile elderly," in *Electronic Government and Electronic Participation, Joint Proceedings of Ongoing Research and Projects of IFIP EGOV and IFIP ePart 2012*, H. J. Scholl, L. S. Flak, M. Janssen, A. Macintosh, C. E. Moe, Øystein Sæbø, E. Tambouris, and M. A. Wimmer, Eds., vol. Schriftenreihe Informatik 39, International Federation for Information Processing. Kristiansand (Norway): Trauner Verlag, Sep. 2012, pp. 215–222.

[6] Øystein Dale, "User needs analysis," MobileSage Consortium, Tech. Rep. MobileSage Deliverable D2.1, 2012. [Online]. Available: http://mobilesage.eu/public-documents/public-deliverables

[7] T. H. Røssvoll, "User requirements specification," MobileSage Deliverable D2.2, MobileSage Consortium, Tech. Rep., Feb. 2012. [Online]. Available: http://mobilesage.eu

[8] L. Curescu, I. Anghelache, and T. H. Røssvoll, "System requirements specification for help-on-demand service," MobileSage Deliverable D2.3, MobileSage Consortium, Tech. Rep., Apr. 2012. [Online]. Available: http://mobilesage.eu

[9] T. H. Røssvoll and V. A. Gracia, "System requirements specification for content management service," MobileSage Deliverable D2.4, MobileSage Consortium, Tech. Rep., Apr. 2012. [Online]. Available: http://mobilesage.eu

[10] R. Mohamad Rasli, S. C. Haw, and R. Mohamad Rasli, "A survey on optimizing image, video, and audio query retrieval in multimedia databases," *International Journal of Advanced Computer Science*, vol. 2, no. 6, 2012.

[11] D. Das and A. F. Martins, "A survey on automatic text summarization," *Literature Survey for the Language and Statistics II course at CMU*, vol. 4, pp. 192–195, 2007.

[12] D. Brezeale and D. J. Cook, "Automatic video classification: A survey of the literature," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 38, no. 3, pp. 416–430, 2008.

[13] W. Hu, N. Xie, L. Li, X. Zeng, and S. Maybank, "A survey on visual content-based video indexing and retrieval," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 41, no. 6, pp. 797–819, 2011.