Web 2.0 Vision for the Blind

Robert Baumgartner
Lixto Software GmbH
Vienna, Austria
baumgartner@lixto.com

Ruslan R. Fayzrakhmanov
Wolfgang Holzinger
Bernhard Krüpl
Max C. Göbel
David Klein
Inst. of Information Systems
University of Technology
Vienna, Austria
abba@dbai.tuwien.ac.at

Rafael Gattringer
Faculty of Psychology
University of Vienna
Austria

ABSTRACT

In today's world, access to the Web has turned into a basic need. The Web eases daily life significantly and has turned into an essential - and in some scenarios even exclusive - channel for information, communication and transactions. However, the Web does not simplify life for everyone. To guarantee the benefit of the Web for the entire human population, the needs of user groups with special requirements such as blind users are of supreme importance. Recently, especially with the advent of Web 2.0 technologies, new barriers have been created.

We believe, the only solution is that a blind person consuming the Web is put back in full charge of how to consume it. In the ABBA (Advanced Barrier-free Browser Accessibility) project described in this paper, on the one hand we create empirical field-studies where we investigate the problems blind users face in the Web, and on the other hand a methodology and system that provides them with the kind of help a sighted assistant would give. To create such a methodology, understanding of how blind users perceive Web pages is essential and reflected by creating mental models. Based on these models, strategies for different axes of navigating a Web page are devised.

Keywords

accessibility, blind, navigation, screen reader

1. INTRODUCTION

For the blind the Internet is still a valley of "blood, sweat and tears", as one of our interview partners describes it. This leads to a feeling of frustration and total exclusion as a potential target group. Yet the advantages of the Net outweigh the induced frequent frustration.

Worldwide knowledge and communication are the main driving forces for enduring the challenges of the Web. Its emergence was in fact a revolution for the blind, perhaps even more important than for sighted people. Available documents for example do not have to be translated any more to Braille but can be directly accessed via the Net. Podcasts give a great example of auditory content.

Web Science Conf. 2010, April 26-27, 2010, Raleigh, NC, USA.
.

Following factors influence the blind Web surfing experience: the structural quality of the Web page, the skill level and motivation of the Web designer, the knowledge of the user and the type of the screen reader.a

Blind users perceive a web page as a stream of single HTML tags flowing by. On the other hand a very broad view of a document is given to sighted users by visual elements. Not only the objects themselves are presented, but also different colours and shapes with their relevance to the psychological based gestalt laws. As thereby much more information is processed at once, the visual channel can be compared to a broadband connection. This latter access channel enables a much quicker exploration of new web sites. When blind users exactly know the document structure of a web site, they are able to navigate even faster than sighted users. For blind users any visual information is lost from the visual driven Web. This results in a completely different user experience.

As technologies shift and new ones emerge, blind users have to constantly adapt their access strategies. When search engines became the dominant portal for accessing the Web, this had a major impact on the quality of the blind user experience. Nowadays competing web sites are forced to optimise their HTML code and site structure to achieve a higher search ranking. In a certain way a search engine crawler extracts similar information like a typical blind user. Both have to strongly rely on HTML markup tags to get an idea of the content of the page in an acceptable time.

On the other hand, many of the developments of the so called Web 2.0 revolution have left the blind community in the cold. When the Web experience became more community-oriented and content creation more decentralised for the average netizen, the sometimes awkward and adhoc realisations based around purely visual paradigms — asynchronous in execution and shrouding content by programming languages — proved to be a formidable obstacle for screen reading software and therefore also for the blind user. As more and more content moves into these networks, and the blind users stay locked out, the quality of their Web experience decreases.

More recently, the tide seems to turn again. Efforts to incubate a Semantic Web, originally intended to help machines grasp a glimpse of comprehension of the content they are processing, might in fact turn out to be a direct profit for the blind community and accessibility technology.

The remainder of the paper is structured as follows: Section 2 continues presenting a broader overview of our experiences and communications with people form the blind community, and the lessons we were able to draw from it. Section 3 introduces the abstract conceptual model of our approach, the mental model and its connection to Psychology's gestalt theory. Section 4 presents a more concrete, formal and implementation–oriented approach to the guidelines laid out in Section 3 and introduces the Multi-axial navigation model. Section 5 is committed to list related work and describes the context of our ideas. Finally, Section 6 concludes the paper and give a further outlook on our future plans.

2. BRIDGING THE GAP TO THE BLIND WORLD

Questions arise when starting the research into the blind world. Do we know if our concepts about blind people are correct? Where have our assumptions originated? How should we communicate to avoid unintended insulting remarks?

We noticed that creating new technology for helping the so called "disabled" lead to a position of unaware superiority. This attitude creates an intention of knowing what is useful and practical for blind people, however not involving them in the development. The mentioned "disabled" view might mislead promising research. In our opinion society forms an environment that either "enables" or "disables" its stakeholders. Just imagine for a moment that a vast majority of the people in the global village were blind. In this scenario we might have mostly advanced (auditive and kinaesthetic) screen readers today instead of highly visually reliant web browsers, resulting in enabled blind people and disabled seeing people.

Therefore we suggest that it is indispensable to start any research from scratch and questioning previous mental concepts. Usability studies can create an insightful understanding for the blind experience.

An interview partner pointed out that the perception of blind people was not worse, just different. We would perhaps call it "shifted perception".

Research methodology. As mentioned, attaining a clear understanding about the blind mental experience seems important. Therefore combined qualitative interviews with usability studies were used to observe and explore this different web user experience. Thinking aloud is a very helpful technique to retrace the perception of the blind users. A questionnaire regarding the usage of the Web was set up as well. The results of both qualitative and quantitative research techniques however have to be taken into account only very carefully, as the sample size of the probands in both approaches is rather small.

2.1 User behaviour

There are two ways of approaching accessibility for blind users: text-to-speech (TTS) systems and Braille displays. These methods are not competing, in fact they could be seen as complementary. The advantage of this combination is that different interpretations of data which provide more specific information to the user. While TTS works well to cover the presentation of occurring events such as pop ups or important notifications and therefore is often used in this context, the unbeaten benefit of Braille is to present more complex contents such as tables, mathematic formulas or similar contents.

Confronted with navigation uncertainty, blind users have a preference for safe havens, consistency and a minimisation of risky navigation choices. One indication for the preference for safe havens might be the strong use of email. The email format symbolises a simple navigation interface and a stripped down layout of the documents. Web forums are replaced by classic mailing lists, confirmations of online banking are sent as email and news might be preferred to be read as mail as well.

Concrete user patterns were especially interesting to observe. Wikipedia for instance is a very frequently visited site of the blind community. The site structure is stable and makes good use of accessibility standards. Wikipedia is often even preferred over common search engines as a primary information source.

2.2 Shortcomings in today's approaches

Screen-readers. The biggest shortcoming of contemporary screen reading software is, in our opinion, its lack of deeper understanding of the large–scale structure of a document.

Existing software, with its roots in general accessibility for graphical user interfaces, treats semi–structured documents as mostly annotated text. It leverages on the presence of obvious markups, like hyperlinks and headings, but makes no further effort in understanding the basic structure of a document. In a way, it is missing the bigger picture while presenting the tiny details.

The problem with this approach is that the user is often overwhelmed with an avalanche of mind–numbing detail, because the system is incapable of distinguishing content from control elements or pure decoration.

Current software more or less strips a Web document of all non–text elements it does not understand and presents the remaining text content.

Further, general drawbacks of some screen-readers are a high purchase price, indirect costs through the forced purchase of an operating system and finally closed source. Especially, free and open source software would be helpful in his area to exchange ideas and enable the blind and seeing community to contribute.

ARIA Roles. The WAI–ARIA [9] initiative addresses most of these problems by proposing a framework of annotations to be put alongside the Web document. These annotations describe the "role" that certain elements of a Web page have with a well–defined vocabulary.

We see two main problems with this approach: (a) inclusion of ARIA markup depends solely on the content creator, who is often too uniformed, works under time constraints or is too under–budgeted to care; and (b) this is still a local approach that, while describing the purpose of single elements in a much more satisfactory way, does not help the user with all around navigation and familiarisation of the page structure. Today, such guidelines are at best applied in governmental sites.

3. BUILDING MENTAL MODELS

Much information is currently inaccessible for blind users, because it is only visually available. It is quite common for web site creators to think in visual terms when they design Web pages, and often they start from visual designs and use visual tools for their work. The navigability of a web site for blind users is often not part of the original design process, but, if at all, only concerned at a later step when accessibility standards are implanted into the otherwise finished site. The visual appearance of a web page almost always stays the golden standard for understanding its logical structure. Regrettably, it is also unaccessible for blind users, making them dependent on web site authors and their proper implementation of accessibility standards. When we talk about visual appearance, we do not focus on images or rich media that are embedded on web pages; we rather concentrate on the general look of a Web page, where the layout, i.e. how text on the page is positioned in which format, plays an important role. Researchers of the first part of the last century described these visual properties with the *gestalt theory,* and we have implemented a system that leverages this view.

3.1 Reverse authoring process

Core to our framework is the idea to make the gestalt information of web pages available to blind users. This helps to improve the navigability of the document, i.e. makes it faster and more convenient for blind users to access those portions of the page that are of interest to them. Using the term gestalt information, we consider all layout information that is used to make the relation of items on a page more clear to sighted users, such as *i) groupings, ii) alignments, iii) common formatting properties,* etc. A number of gestalt principles have been postulated that help a user to identify which objects in a larger set belong together, e.g. because they are close to each other or look the same. A discussion of these principles is out of the scope of this paper, but we have been investigating how they are applicable to web site design. Also, we created a gestalt ontology that captures all important gestalt properties of web page items and can describe a web page in using these gestalt concepts. This essentially reverses the authoring process of a web page to extract exactly those concepts that were at the root of the design process. An examples for a gestalt analysis of a Wbb page is the identification of visually outstanding or salient text: Any text items that are rendered larger and bolder than most other text items are potential headings, even if they are not properly annotated as such in the original HTML source code.

On the other hand, we want to emphasise that while gestalt information can be highly valuable in the understanding of a Web page, we believe that blind users should not be confronted with a 1:1 transcoding of a visual interface. Interfaces such as VoiceOver for Safari on the iPhone—where users can tap on any part of the screen to get a voice description of the content at this location—might be fascinating, but interviews with our users reveal that working with this 2-dimensional environment is unintuitive to many blind users. They see no use in identifying visual objects on pages per se. In contrast, our approach is to augment the stream based surfing experience of blind people with hints that stem from visual properties in order to identify those visual objects that can be associated with certain roles.

Web design patterns. There are several frameworks that try to identify subregions on web pages, and ours is not different in this respect. In addition to that, we can see that web authors often make use of web design patterns. A number of public web design pattern libraries have been published that should help designers in picking the right pattern. We can see patterns for navigation, searching, making choices, shopping, etc.; examples for patterns include breadcrumb navigation, carrousels for data display and date pickers. They reflect a common understanding of a function and became abundant when the Web moved from a collection of documents to a more complex application interface.

Patterns usually consist of a number web page elements, sometimes grouping also non-contiguous parts of a page or even parts of several different pages. Starting with a subset of these patterns, it is a goal for the next future of our accessibility framework to automatically detect and annotate them by using rules and heuristics. As good web design usually incorporates these patterns, our system will be able to deal better with well-designed and also visually appealing web pages, which in fact are usually also easier and faster to understand for sighted users.

3.2 Building mental models

Why do we need to talk about mental models for web pages at all? Because the reading strategies of sighted and blind users differ: Whereas sighted users can just look at a page to get an idea about its structure, blind users have to build a mental model of the page. Whether they depend on a text-to-speech or braille interface, blind users need to parse the serialised data that comes across the interface into their mental model for later reference, thus enabling them to get an overview of the page.

Requirements for the mental model. The mental model for a web page needs to support two basic *scenarios: i) visiting a page for the first time,* and *ii) revisiting a page.* The first scenario demands for a model that can quickly provide a good overview, by collecting different aspects the page into a common model. In the second case the focus is on enabling fast access by efficiently navigating along known landmarks. In addition to these scenarios, two different reading *modes* should be supported: In what we call the *hunter mode,* the user is on the search for some particular information on the page, e.g. the opening hours of a restaurant. In contrast, in the *tourist mode,* the user has no particular goal and rather wants to get an overview of the page.

What do the mental models of blind users look like? Takagi [23] proposed an interesting view of a mental model after performing an empirical study with five blind probands. We believe, though, that there is actually no mental model that is universal to all Web pages and blind users; the mental model of a user is rather dependent of: *i) the Web page* under consideration, and *ii) the screen reading environment,* which consists of the combination of all necessary hardware and software parts. We further assume that by carefully selecting its user interface, a Web screen reader can support and thus determine a specific mental model in its user—that, within certain constraints, we are free to choose, and thus should be as inspiring as possible.

On the other hand, we must strive not to over-engineer the model. It is quite clear that it should be richer than just a text string, but it is important that it will not put too much mental workload on the user. Also, even a modern blind Web browsing environment should not try to be too intelligent on its own but rather enable the intelligent user to comfortably access all required information.

Reading patterns. Blind users can employ a variety of different navigation strategies or reading patterns, and these strategies should be well supported by the mental model.

Local reading patterns are patterns that can be accessed at any point while reading a document:

- *Play.* The most basic pattern starts voice output from the current cursor position. Once started, information is pushed to the user in a pre-defined speed until stopped.
- *Read.* When using a braille line, the user can control how fast they want to read text (pull).
- *Skip-all-links* is a concept that is built-in into modern screen readers; other similar concepts exist.

Global reading patterns are those patterns that characterise the principal approach of how to navigate through a web page:

- *Text-search* (typically Ctrl-F). Full-text search is very important to blind users, and they often try to guess words that could appear close to the relevant content they are looking for.
- *Web site-search.* Since many modern Web sites feature a site-wide search functionality, as a first step users often try to locate this functionality and use it to locate relevant content.
- *Hierarchy.* A hierarchical reading pattern parses a page into fragments that are leafs of a page document tree. This is quite natural to many web developers, since it reflects the internal DOM structure of a web browser, but it can be argued that it is not that natural to the average user. In other approaches, the visual rendition can be used for segmentation. In our experience, blind users face difficulties when accessing pages in this hierarchical manner, because their mental model of web pages usually is not of visual nature.
- *Stream-oriented.* We observed that blind users find a stream- or string-oriented navigation often liberating, because it needs less mental workload. The most straight-forward model for one-dimensional information is a string of text, and we all are used to this representation when we read a book or listen to the news on the radio. The problem with streams is that they are hard to navigate because there is no structure.
- *List-oriented* = multi-axial stream access. Most current screen readers follow this paradigm: Web pages are represented as a one-dimensional stream (see above), and additional index lists provide access to certain points (landmarks) in the stream. Examples for such indexes include lists for HTML heading elements ($\langle h1 \rangle$, $\langle h2 \rangle$ etc.) and link elements ($\langle a \rangle$ tags), but one can imagine additional indexes that are not based on the HTML source.

ABBA's mental model. We have designed ABBA to support a mental model for multi-axial scanning navigation on Web pages. Keeping the user oriented is one of our premier concerns, and therefore we base our model on a strictly one-dimensional stream representation of the page. In [23], the authors observed one-dimensional mental model for Web pages that is divided into several fragment regions, where fragments correspond to areas on the page, e.g. the main content area. The DANTE system [25] also marks fragments on a page and divides it physically into a number of smaller and simpler pages, or alternatively enables a better logical access by creating a table of contents to the top of the page. We believe that there is not just a single segmentation algorithm that universally can divide Web pages into meaningful fragments, and that the segmentation should depend on the task a user tries to achieve when accessing—the perspective on the page will change when changing the task. In a way, current screen readers allow switching the perspective with their list-oriented access modes. Users can decide to jump from HTML heading to HTML heading, or they can with to another mode where they can navigate from HTML link to HTML link. Our system enhances this already multidimensional reading patterns by introducing additional navigation paths that are based on the gestalt view of the page (e.g. accessing all visual headings of a page). Our model is actually multi-axial, but we realized that in order to avoid cognitive overload of the reader it is better to hide the multiaxial aspect behind a strictly one-dimensional model, where i) all text is arranged in a linear way, and ii) landmarks of different types are placed on the text. Users can quickly access these landmarks by selecting a certain list of them and cycling through them. Mentally, this type of access is closely related to a scanning navigation where users scan for rich objects (such as headings or articles).

Canonical reading order. Even when providing multiaxial access to the page, we need a reference order in which the items (words / lines) of a page are serialised. This is because of the *multiple segmentation problem:* Different perspectives on a page require different segmentations—a news page can be divided into columns or articles, and fragments of these respective types could even overlap. In consequence, different segmentations of a page lead to different serialisations. If users are confronted with a different ordering of a page whenever they change their perspective (e.g. when switching from column mode to article mode), they would be totally lost. Therefore, ABBA needs to pick one particular segmentation and reading order upfront (the canonical reading order) and stick to it. The different types of segmentation have to be supported by supplying pointers to landmarks of different types that mark the beginning of fragments. In selecting a suitable canonical reading order, we see several options: *i) source order,* this chooses the order of text elements as they appear in the HTML source; *ii) DOM order,* which loads a web page, applies all Javascript and dumps the resulting DOM tree; *iii) segment order,* which runs a segmentation algorithm on the web page and sorts and serialises the resulting segments. ABBA uses the third variant, and its segmentation algorithm is based on a the browser-based rendition of the web page. At any time when a user access a Web page, they can ask for their position in the document, and they will get back an indication of at how many percent of the serialised document the virtual cursor is currently located—based on the canonical reading order.

4. ABBA METHODOLOGY

Navigation on the Web typically boils down to user interactions within the browser. While this is acceptable on the click-and-drag paradigm of modern graphical user interfaces, navigation interactions are very expensive for the blind user that does not have access to a mouse-centred interface. It must be the goal of accessibility research to reduce

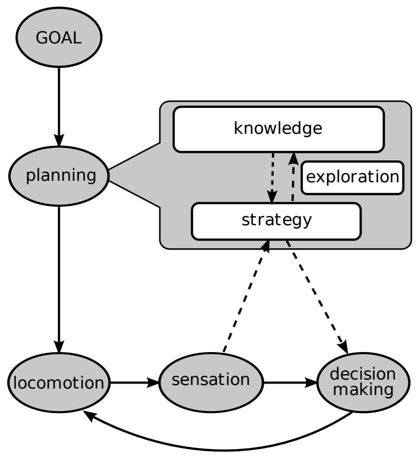

Figure 1: The cognitive navigation model

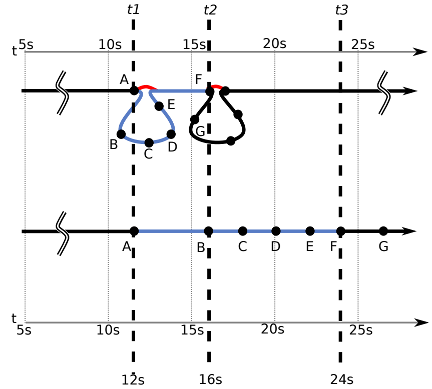

Figure 2: Time comparison of different document serialisations. Top: Time-optimised folding on recurrent patterns. Bottom: Full document serialisation

the number of interactions and therefore to simplify navigation as much as possible. We approach this problem from a cognitive perspective, by means of *Information Foraging Theory* (IF) [20].

4.1 Understanding navigation models

Cognitive navigation. Cognitive research describes navigation as follows [12]: An agent is located in a world (its environment) and is given a *goal* to achieve. Based on some existing a-priori knowledge about both the domain of the goal and the agent's environment, the intelligent agent would then proceed to construct a *strategy* of how to best achieve its goal. This is also known as the *planning stage*. A possible lack of knowledge needs compensating via *exploration*. For complex scenarios a goal is often broken down into smaller sub-tasks. When the agent has arrived at a plan, it starts executing it step-wise through *locomotion* (locomotion here not necessarily needs to be of physical nature, but should rather be seen as a state transition of an automaton). The locomotion step returns a perceptual feedback (sensation) to the agent which is used for orientation. In case of unexpected feedback the agent might update its knowledge of the world accordingly. Based on its strategy the agent now alternates decision-making steps with locomotion steps until the goal is achieved. The concept of navigation is summarised in Figure 1.

In a classical cognitive sense, *orientation* is the spatial understanding of the agent's whereabouts in its world respective to a set of landmark objects. In the real world, we typically use maps together with salient objects close by (referred to as *landmarks*) to determine our exact position. If we lose orientation we resort to an explorative strategy until we regain orientation. Orientation therefore plays a key role in the decision-making step.

Given the above definition of navigation as the sequence of locomotions to reach a goal allows us to now describe its effectiveness. IF research studies information-gathering and sense-making strategies from a psychologic perspective. The hypothesis therein is that a natural information system will eventually approach a stable state where the chosen strategy maximises the gains of valuable information per unit cost [20]. We can therefore measure the effectiveness of a navigation as the cost of the locomotion with respect to reaching the goal. Thereby we can also associate a precise value with each locomotion as the fraction that locomotion takes us closer to the goal multiplied by the value of the goal. One of the core concepts behind sense-making are so-called *Information Scents*— subjective costs of accessing and retrieving information bits based on perceptive cues [19, 8, 7, 5, 21, 24, 18].

Comparison of navigation strategies. Access to information on the Web typically requires interaction by the user to navigate to the information she requires. We can distinguish between two types of document navigations: inter–document navigation, describing travel between several documents; and in–document navigations, describing travel between regions of a single document. In this research we are interested in the navigation within a document.

A typical user strategy is to minimise the time spent to retrieve some particular information and thus to make optimise navigation. For document reading, this is achieved by defining so-called *skip-points* that prune away content of little to no informational value. In Figure 2 we describe how skip-points can be understood in terms of a document serialisation. The figure compares a fully expanded document serialisation at the bottom with a collapsed serialisation of the same document at the top. Collapsing items is referred to as *folding* and it is done on items that act as descriptive containers of their children (e.g. news headers). On the x-axis we display the time in seconds needed for each serialisation. The figure therefore shows the speed-up (16 vs. 24 seconds to reach point F) a reader can achieve through the intelligent use of skip-points. Skip points in our understanding of Web navigation can be obtained directly through the information scents of a document.

4.2 Multi-axial navigation model

Model definition. The *Multi-axial navigation model* is a set of predefined axis types and axes applied to the model which describes particular domain, and also any relations between axes. Thus, our model is built on top of what we call the Unified Ontological Model [10]. This model contains information about rendered web document with its visual representation, and it is generated from web page analysis and applying Web data extraction algorithms. For creating the Multi-axial model, the underlying model should de-

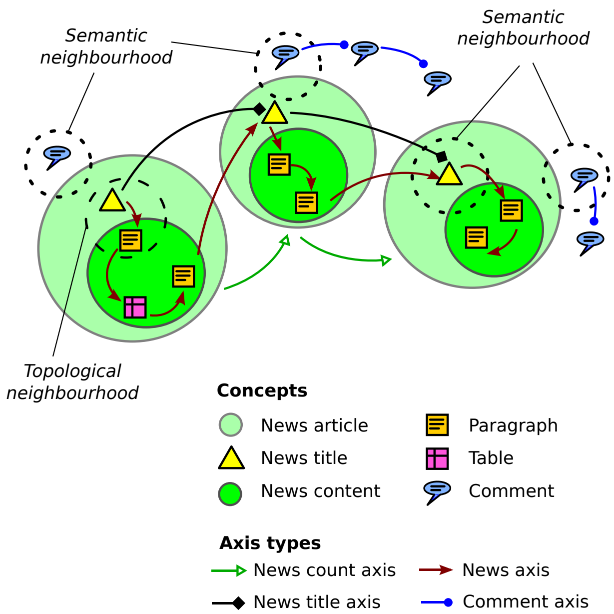

Figure 3: Example of the Multi-axial navigation model for a news web site

fine objects with topological and semantic relations between them, and visual features such as font, colour, size and so on.

An *axis type* represents the abstract rule for reading certain concepts of the underlying model and thus defines a rule for ordering particular concepts. An *axis* is an instance of an axis type. It is created by applying rules of ordering for the Unified Ontological Model. So, axis is simply ordered list of objects.

In Figure 3 we give an abstract example of navigational axes within a hypothetical news site. The depicted model represents main concepts of the news domain such as "news article", "news title", "content", "comment" etc. These concepts can have a complex structure on their own and include further sub–concepts, e.g. news articles that can contain sections, paragraphs, tables, figures etc. Depicted concepts with defined topological and semantic neighbourhoods are part of the Unified Ontological Model. In the figure, the model also represents four types of axes: "news title axis", "news axis", "comment axis", and "news count axis"". Axes types applied for the model can have one instance (e.g. "news title axis") or several (e.g. "comment axis"), or not at all. Axis can have arbitrary quantity of elements.

There are two types of relations which define the neighbourhood in the Multi-axial model: topological and semantic. Topological relation is based on relative distance between objects and defined in the Unified Ontological Model. The example can be the topological relation between news title and a paragraph which is the first paragraph of the news article (Figure 3). Semantic relations can be set either for objects in the underlying model or for axis types in the Multi-axial model. For the semantic relation defined between the axis types the semantically nearest object of the particular axis will be the object of another axis which is topologically nearer than other ones of the same axis. For instance, relation between "news axis" and "comment axis", and relation between "news title" and "news content" as an example of semantic relations. Any relation in our model is the quadruple $\langle o_i, o_j, \rho_k, w_{i,j,k} \rangle$, where o_i and o_j are objects of the models, ρ_k — the type of relation, $w_{i,j,k}$ — weight. For the topological relation the weight is a distance between objects. Relations of the same type with the certain value of length of transitive chains for the particular object represents its semantic or topological neighbourhood.

Model Navigation. According to IF theory we introduce respective concepts for our Multi-axial model. We see two different types of locomotions for our model: *in-axis navigation* and *inter-axis navigation*. Basic in-axis navigation has following operations: i) *next* (go to next object of the current axis), ii) *previous* (go to previous object of the current axis), iii) *first* (go to first object of the current axis), and iv) *last* (go to last object of the current axis). Basic operations for inter-axis navigation are v) *set new axis* (go to the first object of the new axis) and vi) *change axis* (go to topologically nearest object of new axis to the current one). All navigational paths are recorded and the user always has the opportunity to go through objects in reverse order.

After the model evaluation and user behaviour analysis we found additional useful operations: vii) go to previous object, where last axes change occurred, and change the axis to the previous one; viii) go to previous object, which is in the neighbourhood of previous axis, and change the axis to the previous one. The first command is used to rollback locomotions which were done at the selected axis to the initial state. The second command is used to always have mapping between an object of current axis and an object of the previous axis. For instance, if the user went through news titles at the news title axis and then changed the axis to the news axis, then he will be able to go to the title of current news article and go through news titles again to find other relevant news.

The inter-axis navigation — possibility to change axis at the particular object, in its topological or semantic neighbourhood — is one of the benefits of our navigation model. The reason for changing axis can be because of finding an item with information scent and the user wants to change the axis from more general to a more specific one. Changing from more specific axis to less specific one can be done because of very low information scent. The changes axes in either topological or semantic neighbourhood tells us about intention to find necessary information serialised in different way, in different axis.

For the orientation inside the axis we use such characteristics as quantity of passed objects and residuary objects. These characteristics can be conveniently represented as a percentage ratio. Main operations on the model for the inter-axial orientation are: i) get list of all axes in the model, ii) get list of all axes intersecting the object, iii) get list of axes in the topological neighbourhood, iv) get list of axes in the semantical neighbourhood.

Inter-axis navigation is an important new concept. It allows the user to explore the structure of a document quickly, with different contexts and goals, and without losing orientation. At any point, if he goes astray, *backtracking* to a more familiar location (landmark) is possible, because the axis grid is a stable network. This principle of "no surprises" encourages users to explore freely and without risk of getting lost, while the coarse grain nature of some of the axes also make it a less time-consuming and therefore less painful experience.

The navigation model presented so far is implemented as a prototype. Axes, where possible, are created automatically, but the systems also allows for manual annotation and axis generation. Web pages are rendered by a web browser, and the resulting visualised document is converted to an RDF representation, landmark recognition and axis generation works on this graph model. The resulting model and navigation on the Multi-axial model is then exposed via either a proprietary user interface and speech output from a command line application or via a Web interface using a conventional screen reader.

5. RELATED WORK

The work presented herein is of two-fold nature: we propose to combine efforts from the cognitive research in measuring accessibility cost of Web documents with the application-oriented research in intelligent Web screen readers.

One research strand in Web accessibility has studied the field of information foraging and information scent [20, 14, 13, 12] as a means to alert the blind user of poorly encoded Web resources in terms of accessibility guidelines.

[11] propose to use travel objects as special landmarks for measuring the navigability of Web documents. [23] take a similar approach to include information scent in the evaluation of accessibility scores of Web documents. Both groups use the information to evaluate a Web document for an accessibility score. They annotate hyperlinks on a document with this score to warn the blind user of possible accessibility obstacles to be encountered when following this link. While this may be a useful thing to do, we use the information scent approach in a quite reverse sense. We generate from a Web page available landmarks or scents and provide these to the user as special landmarks for a better understanding of the page domain.

As our goal is to improve the Web experience of the blind user, our work also compares to existing "intelligent" screen readers. There have been a number of screen reading systems in particular in recent years.

Most closely related to our work is the HearSay screen reader [22, 4, 3]. The authors perform a structural and semantic segmentation of the HTML source (DOM tree) and generate a semantic partition tree for a document. An XML-based audio dialog manager provides the navigation and speech output interface to the blind user. HearSay also uses geometric clustering as one part of its segmentation algorithm. CSurf [17] is an extension of the HearSay screen reader by a context analysis model. The authors propose to use a statistical learning method (SVM) for ranking free text blocks in a document into context multi-sets by relevance. In a separate step, they load the corresponding detail pages if available and map the relevant text onto the summary context block of the initial page. The biggest difference to ABBA is that we follow a more cognitive approach to Web document navigation, where we support the blind user in generating a solid mental model of a document through multi-axial landmarks.

Visual elements play an important role in Web page navigation. A recent eye-tracking study undertaken at the University of Manchester [26] shows that sighted users make excessive use of what the authors call Way Edges—visual constructs that group content into sections. We believe that such constructs can be more easily retrieved using a unified data model as we suggest in our framework.

WebAnywhere [2] is a screen reading tool that is published as a free Web service in contrast to a desktop application. The obvious advantage is the ubiquitous access via the Web without the need of a preinstalled software. The system uses a depth-first search on the DOM tree of an HTML page to give the blind user better control over the Web content, and streams back the audio output in MP3 format. WebAnywhere's focus is on the distributed packaging over the navigability of the system.

TrailBlazer [1] and AxsJAX [6] study the navigation behaviour of the blind user by leaving trails along the path taken by the user across a set of Web documents. In particular, AxsJAX uses Content Navigation Rules (CNR) to explicitly define navigation trails through the construction of lists of items. These lists are quite similar to ABBA's axes, but CNRs have to be compiled manually and are based on XPath rather than a unified model that also considers rendition information.

SADIe [15, 16] is a transcoding proxy that uses semantic annotations of CSS style sheets in order to improve Web access for blind users. It relies on the manual construction of website ontologies that extend an upper ontology for the identification of key elements of a page and generation of CNR for the websites. SADIe leverages the fact that Web designers often create specific CSS elements for content blocks that share a role, but is also dependent on that HTML blocks can at least be distinguished using CSS selectors.

6. CONCLUSION

The Web is constantly evolving towards a social platform and the creation of personalised views. Both aspects are crucial in the removal of barriers. Besides creating a better understanding for each other, the community aspect helps due to users that annotate Web pages and hence extend existing navigation axes. Additionally, personalised views are of heavy concern for other user groups, such as elderly people.

The developed technologies pave the way to other application areas as well, such as to repackage content for various end devices. In particular, small mobile devices require an efficient way to represent the interaction semantics and relevant content to the user on a small screen and with very limited interaction possibilities. Based on the representation and interaction model generated for each web page, APIs for mashup assemblers can be provided. And finally, in B2B scenarios such as Online Market Intelligence the developed methodologies help to understand web process flows and content structures.

The blind perhaps provide the "ultimate quality assurance" as semantic markup is indispensable for an acceptable web surfing experience. We strongly recommend any web designer to test own websites with blind users! All of our blind probands were very interested in contributing to research. One blind lecturer mentioned that it was of special concern for him to get the message out that blind people are also contributing researchers.

The ABBA project wants to help to open the gate to today's World Wide Web, in particular to Web 2.0, to blind and partially sighted users. Screen readers can build on the ABBA model to access the large amount of Web pages that do not conform to accessibility standards — not only to present information, but also to guide users to the relevant pieces of information and offer succinct user interaction steps.

7. ACKNOWLEDGMENTS

The ABBA project (Advanced Barrier-free Browser Accessibility) is sponsored by the Austrian Forschungsförderungsgesellschaft FFG under grant 819563.

We would like to thank Jürgen Schwingshandl and Wolfgang Nowak for providing valuable discussions as well as providing us with contacts within the local blind community.

8. REFERENCES

[1] J. Bigham, T. Lau, and J. Nichols. Trailblazer: enabling blind users to blaze trails through the web. In *Proceedings of the 13th international conference on Intelligent user interfaces*, volume 09. ACM, 2009.

[2] J. Bigham, C. Prince, and R. Ladner. WebAnywhere: a screen reader on-the-go. In *Proceedings of the 2008 international cross-disciplinary conference on Web accessibility (W4A)*, pages 73–82. ACM New York, NY, USA, 2008.

[3] Y. Borodin, J. P. Bigham, A. Stent, and I. V. Ramakrishnan. Towards one world web with HearSay3. In *Proceedings of the 2008 international cross-disciplinary workshop on Web accessibility (W4A) - W4A '08*, pages 130–131, New York, New York, USA, 2008. ACM Press.

[4] Y. Borodin, J. Mahmud, and I. Ramakrishnan. The HearSay Non-Visual Web Browser. (Vxml):128–129, 2007.

[5] S. K. Card, P. Pirolli, M. V. D. Wege, J. B. Morrison, R. W. Reeder, P. K. Schraedley, J. Boshart, X. Palo, A. Research, and P. Alto. Information Scent as a Driver of Web Behavior Graphs: Results of a Protocol Analysis Method for Web Usability. (3):498–505, 2001.

[6] C. L. Chen and T. V. Raman. AxsJAX: a talking translation bot using google IM. *Proceedings of the 2008 international cross-disciplinary workshop on Web accessibility (W4A) - W4A '08*, page 54, 2008.

[7] E. Chi, P. Pirolli, K. Chen, and J. Pitkow. Using information scent to model user information needs and actions and the Web. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, page 497. ACM, 2001.

[8] E. Chi, P. Pirolli, and J. Pitkow. The scent of a site: A system for analyzing and predicting information scent, usage, and usability of a Web site. *CHI 2000, Human Factors in Computing Systems. NY: ACM.*, 2000.

[9] J. Craig, M. Cooper, L. Pappas, R. Schwerdtfeger, and L. Seeman. Accessible Rich Internet Applications (WAI-ARIA) 1.0. 2009.

[10] R. Fayzrakhmanov, M. Göbel, W. Holzinger, B. Krüpl, and R. Baumgartner. A Unified Ontology-Based Web Page Model For Improving Accessibility. *WWW 2010,to be published*, 2010.

[11] C. Goble, S. Harper, and R. Stevens. The Travails of Visually Impaired Web Travellers. *Science*, pages 1–10, 2000.

[12] S. Jul and G. Furnas. Navigation in Electronic Worlds (Workshop Report), 1997. *Online in lnternet: URL: http://www. si. umich. edu/˜*, pages 1–7.

[13] M. Kitajima, M. H. Blackmon, and P. G. Polson. Cognitive Architecture for Website Design and Usability Evaluation: Comprehension and Information Scent in Performing by Exploration. 2005.

[14] M. Kitajima, P. G. Polson, and M. H. Blackmon. CoLiDeS and SNIF-ACT: Complementary Models for Searching and Sensemaking on the Web. *Religion*, pages 1–16.

[15] D. Lunn, S. Bechhofer, and S. Harper. A user evaluation of the SADIe transcoder. *ACM SIGACCESS Conference on Computers and Accessibility*, 2008.

[16] D. Lunn, S. Harper, and S. Bechhofer. Combining SADIe and AxsJAX to improve the accessibility of web content. *Proceedings of the 2009 International Cross-Disciplinary Conference on Web Accessibililty (W4A) - W4A '09*, page 75, 2009.

[17] J. Mahmud, Y. Borodin, and I. Ramakrishnan. CSurf: A Context-Driven Non-Visual Web-Browser. 1(c), 2007.

[18] C. Olston and E. Chi. ScentTrails: Integrating browsing and searching on the Web. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 10(3):177–197, 2003.

[19] P. Pirolli. Computational models of information scent-following in a very large browsable text collection. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '97*, pages 3–10, 1997.

[20] P. Pirolli and S. Card. Information foraging. *PSYCHOLOGICAL REVIEW-NEW YORK-*, 106:643–675, 1999.

[21] P. Pirolli, S. K. Card, and M. M. Van Der Wege. The effects of information scent on visual search in the hyperbolic tree browser. *ACM Transactions on Computer-Human Interaction*, 10(1):20–53, March 2003.

[22] I. V. Ramakrishnan, A. Stent, and G. Yang. HearSay: enabling audio browsing on hypertext content. *In WWW* , 04:80–89, 2004.

[23] H. Takagi, S. Saito, K. Fukuda, and C. Asakawa. Analysis of navigability of Web applications for improving blind usability. *ACM Transactions on Computer-Human Interaction*, 14(3):13–25, September 2007.

[24] M. Vigo, B. Leporini, and F. Paternò. Enriching web information scent for blind users. In *Proceeding of the eleventh international ACM SIGACCESS conference on Computers and accessibility*, pages 123–130. ACM, 2009.

[25] Y. Yesilada, S. Harper, C. Goble, and R. Stevens. DANTE : Annotation and Transformation of Web Pages for Visually Impaired Users. In *Proceedings of the 13th International Conference on World Wide Web - WWW '04*, pages 490–491. ACM, 2004.

[26] Y. Yesilada, C. Jay, R. Stevens, and S. Harper. Validating the use and role of visual elements of web pages in navigation with an eye-tracking study. *Proceeding of the 17th international conference on World Wide Web - WWW '08*, page 11, 2008.