

Zur akustischen Präsentation visueller Webinhalte – VoiceXML meets HTML & WML

1 Einleitung

„The power of the Web is in its universality. Access by everyone regardless of disability is an essential aspect.“

— Tim Berners-Lee, W3C Director und Erfinder des World Wide Web

Das World Wide Web ist in erster Linie ein visuell ausgerichtetes Medium. Informationen werden codiert, um auf Displays verschiedener Größe angezeigt zu werden; der Anwender navigiert durch diese Daten mit Hilfe von Eingabegeräten wie Maus oder Tastatur. Nicht jeder Benutzer kann jedoch auf das Internet in dieser Art und Weise (durch *Graphical User Interfaces, GUI*) zugreifen, sei es, weil er aufgrund von körperlichen Beeinträchtigungen eine Maus oder Tastatur nicht bedienen kann oder unter einer Sehbehinderung leidet und somit keine visuell vermittelten Informationen aufnehmen kann, oder weil er in bestimmten Situationen seinen Blick nicht auf ein Display richten kann – wenn er beispielsweise hinter dem Steuer eines Autos sitzt.

Es sollte im Sinne der Autoren von webbasierten Informationen sein, ihre Internetseiten leicht zugänglich (*accessible*) zu machen und dabei viele verschiedene Arten der Ein- und Ausgabe zu unterstützen, um somit einen größtmöglichen Anwenderkreis zu erschließen. Denkbar ist beispielsweise ein rein akustischer Zugriff auf das Internet, der für die Ein- und Ausgabe natürliche gesprochene Sprache vorsieht: Die Informationen werden durch ein Sprachsynthesystem vorgelesen, der Benutzer navigiert durch Sprachbefehle. Derartige *akustische Browser* (auch *Voice Browser* genannt), die diesen Zugriff realisieren, sind insbesondere für die Gruppe der blinden und sehbehinderten Benutzer interessant, aber auch für den mobilen Anwender, der mit solchen Systemen Internetinformationen durch ein normales Telefon abrufen kann.¹ Dafür muss ein Weg gefunden werden, die Informationen des Internets in eine akustische Form zu überführen und diese anschließend so zu präsentieren, dass der Benutzer schnell und intuitiv an die gewünschten Informationen gelangen kann.

Diese Arbeit beschäftigt sich mit der akustischen Präsentation von Informationen aus dem Internet. Die untersuchten Formate sind *HTML* (HTML 1999) als am weites-

¹ Internetfähige Mobiltelefone sind verfügbar, jedoch weitaus weniger verbreitet als herkömmliche Telefone.

ten verbreitete Sprache zur Auszeichnung von Texten im Internet sowie *WML* (*Wireless Markup Language*; WML 2001), die Markupsprache für das ‚mobile Internet‘ auf Endgeräten wie Handy und PDA. Es werden verschiedene Ansätze zur akustischen Präsentation von HTML- und WML-Dokumenten demonstriert sowie auftretende Probleme und Schwierigkeiten diskutiert, u.A. anhand eines konkreten akustischen Browsers, der im Rahmen des Projektes „Retivox“ am Institut für Kommunikationsforschung und Phonetik der Universität Bonn entwickelt wurde.

Anschließend wird *VoiceXML* (Andersson et al. 2001, Abbott 2002, VoiceXML 2002) als ein Formalismus zur Beschreibung von natürlichsprachlichen Mensch-Maschine-Dialogen vorgestellt, der aus der Notwendigkeit einer eigenen Markupsprache für akustische Ausgabemedien entstanden ist. Mit Hilfe dieser Sprache können Dialoganwendungen, die von entsprechenden Voice Browsern umgesetzt werden, plattformunabhängig und ohne großen Aufwand erstellt werden.

HTML- bzw. WML-Seiten können nun nicht mehr nur durch Entwicklung spezieller Browser akustisch aufbereitet werden. Die Arbeit untersucht, inwiefern HTML/WML-Dokumente durch XSLT-Transformationen (XSLT 2002) in das VoiceXML-Format übersetzt und dann durch VoiceXML-Browser interpretiert werden können. Es zeigt sich, dass gerade WML eine Verwandtschaft zu VoiceXML aufweist, die eine beinahe vollautomatische Transformation der Formate gestattet. WML-Seiten werden dadurch ohne großen programmiertechnischen Mehraufwand durch jedes beliebige Telefon abrufbar.

2 Die Lingua Franca des Web: Hypertext Markup Language (HTML)

Mit Hilfe von HTML können Texte hinsichtlich ihrer Struktur und ihres Aufbaus codiert werden. Der Autor hat beispielsweise die Möglichkeit, Textabschnitte als Überschriften oder Absätze zu markieren, aber auch Tabellen und Listen als Strukturierungselemente einzusetzen. Die Sprache wurde ursprünglich für visuelle Ausgabemedien konzipiert; HTML-Dokumente werden i.d.R. verfasst, um auf einem Computerbildschirm dargestellt werden zu können. Für die *akustische* Darstellung dieser Dokumente ergeben sich besondere Probleme, die im Folgenden kurz angerissen werden.

HTML-Seiten werden auf einem Bildschirm räumlich, und damit nicht-linear, dargestellt. Ein Autor macht sich bei der Gestaltung einer Webseite Gedanken über die räumliche Anordnung der Informationen, die vom Rezipienten möglichst schnell erfasst werden sollen. Er kann sich dabei eine wichtige Eigenschaft unserer visuellen Verarbeitung zu Nutze machen, nämlich die, dass i.d.R. ein einziger Blick ausreicht, um den groben Aufbau einer Seite zu erschließen. Diesen Prozess fasst man unter den

Begriff der *Makro-Analyse* (Roth et al. 1998). Hier werden Dokumentstruktur und grundlegende Elementtypen wie Texte, Bilder und Formulare erkannt. Durch bestimmte visuelle Hinweise (wie Fettdruck, Großbuchstaben, Linien, Bilder etc.) kann die Aufteilung der Informationen auf dem Bildschirm noch unterstrichen werden. Dies erleichtert die Orientierung und Navigation auf einer Webseite für den sehenden Benutzer (Goose&Möller 1999, Kennel et al. 1996). Er kann hierdurch schnell zu den gewünschten Stellen auf der Seite springen und dann einzelne Elemente, also etwa Textabschnitte, lesen und verarbeiten (*Mikro-Analyse*).

Bei der akustischen Darstellung einer Webseite fallen die o.g. Hilfestellungen weg. Auditive Daten gelangen auf einem Wahrnehmungskanal zum Anwender, der wesentlich weniger Informationen pro Zeiteinheit übertragen kann als der visuelle Kanal. Der Anwender hat also nicht ohne weiteres die Möglichkeit, in kurzer Zeit die Struktur der Seite zu erfassen oder etwa Texte zu ‚überfliegen‘. Möchte er jedoch auf gleiche Weise wie im Falle der visuellen Darstellung den Aufbau der Webseite begreifen, so muss er sich nicht nur die gesamte Seite vorlesen lassen – er muss sich dabei auch mit einer wesentlich niedrigeren Geschwindigkeit zufrieden geben.²

Für blinde und sehbehinderte Benutzer, die an ihrem Rechner im Internet surfen möchten, existieren diverse Möglichkeiten, die Navigation durch das Netz zu erleichtern. Für Anwender mit einer Sehschwäche können Programme eingesetzt werden, die bestimmte Ausschnitte auf dem Bildschirm vergrößert darstellen. Blinde Personen können auf ein *Braille-Display* ausweichen, mit dessen Hilfe sie die Texte des Bildschirms mit ihren Fingern ertasten können (Hahn 1994). Alternativ kommen *Screen Reader* zum Einsatz, die diese Texte vorlesen.

Meist beschränken sich diese Systeme jedoch auf eine bloße Übertragung der reinen textuellen Informationen in synthetische Sprache; eine adäquate Repräsentation bestimmter grafischer Merkmale (wie Fettdruck, Schriftgröße etc.) fehlt. Gerade diese Merkmale jedoch sind für das Verständnis der Struktur eines Textes oftmals sehr wichtig, so dass die Screen Reader-Technik für die akustische Präsentation von HTML-Dokumenten unzureichend erscheint (Goose&Möller 1999, James 1996, Kopecek 1998, Kennel et al. 1996).

Um dieser Aufgabe gerecht zu werden, muss also eine angemessene Umsetzung der HTML-Struktur erreicht werden. Feldes&Stegmann (2000) sprechen hier von einer notwendigen „Medientransformation von der grafisch-textuellen Präsentation in eine entsprechende akustische Präsentation“ (S. 95). Dies soll aber nicht bedeuten, dass diese Transformation ausgeht vom grafischen Erscheinungsbild auf einem Computerbildschirm, sondern vielmehr von der HTML-Codierung, d.h. von den eingesetzten Tags selbst: „[A]udio renderings can be designed from the markup instead of by gues-

² Unter einer Erhöhung der Vorlesegeschwindigkeit, die selbstverständlich möglich ist, leidet schnell die Verständlichkeit.

sing from visual representation“ (James 1996, S.97f.). So können Überschriften beispielsweise anhand der Tags `<h1>...<h6>` erkannt werden und nicht etwa aufgrund der Schriftgröße. Horiuchi et al. (1999) führen den Begriff der „logischen“ Struktur ein, die gegenüber der „grafischen“ Struktur Ausgangspunkt der Verarbeitung sein soll.

Problematisch in diesem Zusammenhang ist jedoch die Tatsache, dass gewisse HTML-Elemente von zahlreichen Autoren nicht in ihrer ursprünglichen Funktion verwendet werden (Goose&Möller 1999). Auch wenn HTML in erster Linie eine Sprache zur Repräsentation von Dokumentstrukturen ist, kommt es nicht ohne Tags aus, die auf das visuelle Erscheinungsbild auf einem Bildschirm abzielen. So kann man beispielsweise mit Hilfe des ``-Elements Einfluss nehmen auf Schriftfamilie, -größe und -stil. Verwendet man nun dieses Element anstelle von `<h1>`, um eine Überschrift erster Ordnung zu markieren, so kann die Absicht des Autors nicht ohne weiteres rekonstruiert werden. Im umgekehrten Fall setzen Autoren die Elemente `<h1>...<h6>` auch ein, um den gewünschten visuellen Effekt (z.B. große Schrift, Fettdruck) für einen Textabschnitt zu erreichen, ohne die inhaltliche Auszeichnung einer Überschrift beabsichtigt zu haben.

Ein weiteres Beispiel sind Tabellen. Die Aufteilung in Zeilen und Spalten ist nicht nur für die angemessene Darstellung zweidimensionaler Daten von Nutzen; Tabellen können auch für das Layout zweckentfremdet werden (was oft der Fall ist), wenn Texte beispielsweise zentriert, bündig oder mehrspaltig dargestellt werden sollen.

Derlei inkonsequente Nutzung des HTML-Sprachumfangs kann verarbeitende Systeme vor Probleme stellen. Von einem kooperativen Autor, der bei der Entwicklung von HTML-Dokumenten neben dem visuellen auch das akustische Erscheinungsbild stets ‚vor Ohren hat‘, kann i.d.R. nicht ausgegangen werden. Feldes&Stegmann (2000) weisen jedoch darauf hin, dass ein in diesem Sinne kooperativer Autor notwendig ist für eine sinnvolle Verwertung von HTML-Seiten für die akustische Ausgabe; ihr Ansatz beispielsweise hat „nicht den Anspruch, das Surfen auf beliebigen Webseiten zu unterstützen“ (S. 97). Schultheiß&Kunkel (1999) gehen noch weiter und behaupten: „... kein noch so ausgeklügelter Parser und Interpret [wird] beliebige HTML- bzw. XML-Seiten befriedigend für einen Sprachdialog aufbereiten.“

Zusammengefasst besteht also die Aufgabe bei der Entwicklung eines akustischen HTML-Browsers im Wesentlichen darin, auf Makroebene die grobe Struktur und den Aufbau einer Seite akustisch zu repräsentieren sowie dem Benutzer die Orientierung und Navigation zu erleichtern und auf Mikroebene eine adäquate Transformation der einzelnen HTML-Tags zu erreichen und die Intention der Autoren bzgl. diverser Layoutentscheidungen zu rekonstruieren.

Die Langfassung dieser Arbeit beschäftigt sich in Kapitel 3 detailliert mit der Frage, mit welchen (akustischen) Mitteln man welche Ergebnisse erzielen kann. Berücksichtigt wird dabei auch die Möglichkeit, mit Hilfe von speziellen Audio Cascading Style Sheets (CSS 1998) Einfluss auf die akustische Ausgabe von HTML-Seiten zu nehmen.

3 Drahtlos ins Internet: Wireless Markup Language (WML)

Innerhalb der letzten Jahre hat sich die *Wireless Markup Language* als ein weiterer Standard für die Auszeichnung von Texten im Internet entwickelt, der auf andere Endgeräte abzielt als HTML. WML wurde konzipiert, um Informationen aus dem Internet beispielsweise auf WAP-Handys oder PDAs anzeigen zu können, welche relativ niedrige Übertragungsraten aufweisen und mit Displays ausgestattet sind, die nur wenige Textzeilen gleichzeitig darstellen können und eingeschränkte grafische Gestaltungsmöglichkeiten erlauben. Aus diesen Gründen ist es wenig praktikabel, den etablierten HTML-Standard als Markupsprache für den mobilen Internetzugang einzusetzen.

Für die akustische Präsentation von Internetinhalten erweisen sich WML-Dokumente nun als interessante Alternative zu HTML-Dokumenten, da sie aus o.g. Gründen wesentlich weniger umfangreich und einfacher strukturiert sind. Dass dies ein Vorteil für die akustische Ausgabe ist, liegt begründet in der Verwandtschaft der intendierten Ausgabemedien. Die Displays, auf denen WML-Seiten i.d.R. betrachtet werden, sind weitaus kleiner als PC-Monitore. Daraus folgt, dass weniger Informationen ‚gleichzeitig‘ dargestellt werden können; d.h., für die Aufnahme der gleichen Menge an Informationen benötigt der Anwender wesentlich mehr Zeit. Die gleichen Gegebenheiten finden wir bei der akustischen Präsentation der Daten vor: Solange die Menge an Informationen nicht reduziert wird, dauert ihre Aufnahme wesentlich länger (vgl. Abbott 2002, S. 4).

Da dies eher unakzeptabel ist, beschränkt man sich bei WML-Seiten auf die Informationen, die für das mobile Internet am geeignetsten erscheinen. Dies hat neben den beschränkten Displays natürlich auch Gründe, die in der niedrigen Datenübertragungsrate von Handys liegen.

Die Arbeit vergleicht das HTML- und entsprechende WML-Angebot mehrerer Informationsanbieter und diskutiert, inwiefern Inhalte für WAP-Anwendungen eingeschränkt werden müssen.

4 VoiceXML

VoiceXML (Andersson et al. 2001, Abbott 2002, VoiceXML 2003) ist eine Markup-sprache zur Beschreibung von Regelsystemen für natürlichsprachliche Mensch-Maschine-Dialoge, die von speziellen Voice Browsern abgearbeitet werden können. Die Standardisierung von VoiceXML vereinfacht das Programmieren anspruchsvoller telefonbasierter Dialoganwendungen; derzeit liegt die XML-basierte Sprache im Stadium einer W3C *Candidate Recommendation* in der Version 2.0 (01/2003) vor.

Voice Browser werden von sog. *Voice Gateways* zur Verfügung gestellt. Diese Gateways bieten gleichzeitig die Telephonie- und Internethardware, die nötig ist, um Anrufe entgegenzunehmen und zu steuern und VoiceXML-Dokumente über das Internet abzurufen. Für eine Auflistung und Beschreibung derartiger Systeme siehe Andersson et al. (2001).

Der Voice Browser hat im Prinzip die Funktionalität eines herkömmlichen Browsers; er analysiert den Markupcode, parst und interpretiert ihn und bereitet ihn für die ‚Darstellung‘ auf dem Ausgabemedium (=Telefon) auf. Als Server für VoiceXML-Dokumente kann ein regulärer Webserver fungieren. Der Austausch der Dateien zwischen Client und Server geschieht über das übliche HTTP-Protokoll.

Zusätzlich zur Interpretation des Markups muss der Voice Browser die Audioein- und -ausgabe, d.h. die Sprachsynthese und -erkennung, verwalten, so dass sich die Architektur einer VoiceXML-Umgebung wie folgt darstellen lässt:

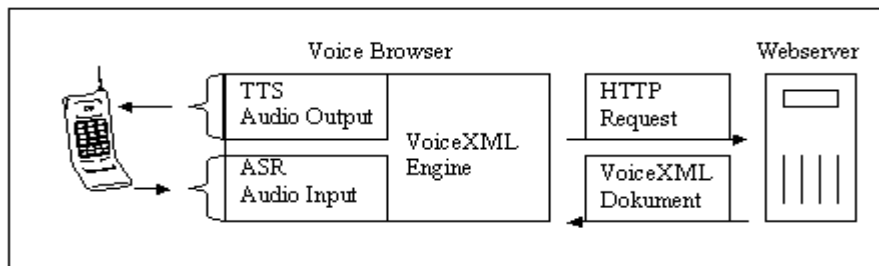


Abb. 1 (aus Andersson et al. 2001, S. 9): Architektur einer VoiceXML-Umgebung.

Ein Beispiel für VoiceXML- sowie WML-Code findet sich im folgenden Kapitel.

5 WML2VXML: Eine Anwendung zur Transformation von WML-Seiten in das VoiceXML-Format

Die Möglichkeiten zur Programmierung von Telefondialoganwendungen, die VoiceXML bietet, können für die Realisierung eines akustischen Zugangs zu Internetinformationen ausgenutzt werden. In Kapitel 7 der Langfassung dieser Arbeit wird ein XSLT-Stylesheet entwickelt, welches WML-Seiten automatisch in das VoiceXML-Format transformiert. Dabei werden Strategien erarbeitet, wie bestimmte WML-Elemente, wie Cards, Eingabefelder, Hyperlinks etc. in VoiceXML repräsentiert werden können. Im Folgenden soll anhand des Hyperlinkkonzeptes in WML der Übersetzungsprozess zwischen beiden Formaten beispielhaft veranschaulicht werden.

5.1 wap.7am.com: Ein Nachrichtenportal

Der englischsprachige Nachrichtenservice 7am.com [letzter Zugriff: 21.1.03] bietet in seinem WAP-Angebot aktuelle Nachrichten aus den Kategorien „World News“, „USA News“, „Sci-Tech News“, „US Politics News“, „Defense and Diplomacy News“, „Business News“ und „Sports News“. Diese Kategorien stehen als Hyperlinks auf der Startseite zur Auswahl bereit und führen zu einer Liste von Schlagzeilen aus dem jeweiligen Bereich, die wiederum mit den Artikeln verlinkt sind. Die Startseite besteht aus einer Card mit dem „title“-Attribut „7am.com WAP News“ und dem einleitenden Text „7am.com News Category:“, wie der folgende Codeausschnitt zeigt:

```
<wml>
  <card id="card1" title="7am.com WAP News">
    <p>
      <b>7am.com News Category:</b><br/>
      <anchor title="World">World News<br/>
        <go href="headlines.cfm?cat=1&amp;t=032359"/>
      </anchor>
      <anchor title="USA">USA News<br/>
        <go href="headlines.cfm?cat=2&amp;t=032359"/>
      </anchor>
      ...
    </p>
  </card>
</wml>
```

Die Kategorien werden hintereinander in <anchor>-Elemente gesetzt, welche Kinder des <p>-Elements sind; dieses wiederum steht unterhalb der einzigen Card dieses WML-Dokuments. Abhängig vom Display des Anzeigegerätes werden diese Daten

unterschiedlich präsentiert. Die folgende Abbildung zeigt das Display eines Siemens-Handys:³



Abb. 2: Die Startseite von wap.7am.com.

Für die Umsetzung der Linkliste in VoiceXML bieten sich mehrere Möglichkeiten an – eine direkte Entsprechung von HTML-/WML-Hyperlinks existiert in VoiceXML nicht. Bei Verwendung eines Menüs (<menu>) beispielsweise dienen die einzelnen Linktexte als Schlüsselwörter (=Sprachkommandos) für die Auswahl einer Option, wie der folgende Codeausschnitt demonstriert:

```
<vxml version="2.0" xml:lang="en-US">
  <form id="card1">
    <block>
      <prompt>
        <voice gender="male">Title:</voice>
        7am.com WAP News. 7am.com News Category:
      </prompt>
      <goto next="#dialog2"/>
    </block>
  </form>
  <menu id="dialog2" dtmf="true">
    <prompt>Please select one of the following options:<enumerate/>
    </prompt>
    <choice next="headlines.cfm?cat=1&amp;t=032359">World News</choice>
    <choice next="headlines.cfm?cat=2&amp;t=032359">USA News</choice>
    ...
  </menu>
</form>
```

³ Unter www.gelon.net [letzter Zugriff: 21.1.03] können WML-Seiten durch einen WAP-Simulator betrachtet werden. Mehrere Handy-Modelle stehen dabei zur Auswahl bereit.

Titel und Text („7am.com WAP News. 7am.com News Category“) werden hintereinander vorgelesen, ebenso die (automatisch generierte) Aufforderung „Please select one of the following options“; das <enumerate/>-Element bewirkt, dass alle Optionen des Menüs ebenso der Reihe nach genannt werden. Durch Sprechen eines der Nachrichtenkategorien bzw. Drücken der durch das „dtmf“-Attribut im <menu>-Element automatisch zugewiesenen Telefontasten 1-7 kann der Benutzer eine Option auswählen.

In der Schlagzeilenübersicht wäre diese Methode jedoch unangemessen, da die Schlagzeilen i.d.R. zu lang sind, um als Sprachkommando zu fungieren. Hier bietet sich eine Umsetzung an, in der die Hyperlinks hintereinander vorgelesen werden und der Benutzer durch einen vorgegebenen Befehl (z.B. „folgen“) den zuletzt genannten Link auswählen kann.

Zahlreiche weitere Beispiele für Transformationen zwischen WML und VoiceXML finden sich in der Langfassung dieser Arbeit.

6 Fazit

Die Arbeit untersucht die Problematik der akustischen HTML- bzw. WML-Browser und zeigt anhand von Beispielen Wege auf, Internetseiten in eine akustische Ausgabe zu überführen. Die Transformation von WML-Dokumenten in das VoiceXML-Format wird beschrieben, mit deren Hilfe WAP-Anwendungen auch für Sprachinterfaces zugänglich gemacht werden können. Da diese Umsetzung weitestgehend automatisch geschehen kann,⁴ die beiden Formate also offenbar eng verwandt sind, wird die Frage aufgeworfen, ob nicht eine Zusammenführung der Sprachen denkbar wäre, etwa eine gemeinsame Markupsprache für mobile Endgeräte, wobei nicht die Mobilität, sondern die beschränkte Darstellungskapazität ausschlaggebend für die Verwandtheit beider Formate ist.

⁴ Das in der Arbeit entwickelte XSLT-Stylesheet funktioniert ohne Kenntnis des Aufbaus bestimmter WML-Seiten. Die auftretenden Schwierigkeiten werden in der Arbeit im Detail diskutiert.

Literatur

- Abbott, Kenneth R. (2002): Voice Enabling Web Applications: VoiceXML and Beyond. Apress.
- Andersson, Eve Astrid et al. (2001): Early Adopter VoiceXML. Wrox Press.
- CSS (1998): World Wide Web Consortium: „Cascading Style Sheets, level 2“. W3C Recommendation.
- Feldes, Stefan/Stegmann, Joachim (2000): Webpage-Call: Ein Gateway zur sprachlichen Interaktion mit Webseiten über Telefon. *Tagungsband Elektronische Sprachsignalverarbeitung (ESSV'00)*, Cottbus.
- Goose, Stuart/Möller, Carsten (1999): A 3D Audio Only Interactive Web Browser: Using Spatialization to Convey Hypermedia Document Structure. *Proceedings of the seventh ACM international conference on Multimedia (Part 1)*, Orlando, Florida, US, 363-371.
- Hahn, E. (1994): Brailleschrift. *Horus – Marburger Beiträge zur Integration blinder und sehbehinderter Menschen*.
- Horiuchi, Yasuo/Atsushi, Fujiwara/Ichikawa, Akira (1999): New WWW Browser for Visually Impaired People Using Interactive Voice Technology. *Proceedings of the 6th European Conference on Speech Communication and Technology (Eurospeech'99)*, Budapest, Ungarn.
- HTML (1999): World Wide Web Consortium: HTML 4.01 Specification. W3C Recommendation.
- James, Frankie (1996): Presenting HTML Structure in Audio: User Satisfaction with Audio Hypertext. *Proceedings of the International Conference on Auditory Display (ICAD'96)*, 97-103.
- Kennel, Andrea/Perrochon, Louis/Darvishi, Alireza (1996): WAB: World Wide Web Access for Blind and Visually Impaired Computer Users. *New Technologies in the Education of the Visually Handicapped*. ACML SIGCAPH Bulletin.
- Kopecek, Ivan (1998): The Architecture of the Blind User Oriented Hypertext AUDIS. *Proceedings of ICCHP 1998*, Wien-Budapest, 215-223.
- Roth, Patrick/Petrucci, Lori Stefano/Assimacopoulos, André/Pun, Thierry (1998): AB-Web: Active audio browser for visually impaired and blind users. *International Conference on Auditory Display (ICAD'98)*, Glasgow, UK.
- Schultheiß, Ulrich/Kunkel, Torsten (1999): Sprachgestütztes automatisches Informationssystem auf Internet-Basis. *Tagungsband Elektronische Sprachsignalverarbeitung (ESSV'99)*, Gölitz.
- VoiceXML (2003): World Wide Web Consortium: Voice Extensible Markup Language (VoiceXML) Version 2.0. W3C Candidate Recommendation.
- WML (2001): Wireless Markup Language version 2 Specification.
<http://www1.wapforum.org/tech/terms.asp?doc=WAP-238-WML-20010911-a.pdf> [letzter Zugriff: 21.1.03].
- XSLT (2002): World Wide Web Consortium: Extensible Stylesheet Language Transformations 2.0. W3C Working Draft.