Daisy 3: A Standard for Accessible Multimedia Books

Dennis Leas
gh, LLC

Emilia Persoon
Dedicon

Neil Soiffer
Design Science

Michael Zacherle
Karlsruhe Institute of Technology

The Daisy standard for multimedia representation of books and other material is designed to facilitate technologies that foster easy navigation and synchronized multimodal presentation for people with print-reading-related disabilities.

People who are blind or have limited vision need alternative versions of books and magazines. Typically, these formats are not available locally and must be specially ordered. They are not as convenient to use as a regular book or magazine because they are large and cumbersome to store and carry around. In addition, printed material causes difficulties for people with some cognitive disabilities, such as those with dyslexia (this group is estimated to be 5 to 17 percent of the population[1]) or motor impairments who can't hold a book or turn its pages. Electronic documents hold the promise of substantially increasing access to material for all of these groups.

For the blind, electronic documents are always available and can be accessed with a portable electronic device. For those with vision-related cognitive disabilities, electronic documents can be presented in a simplified layout, with special fonts or audio synchronized with word highlighting to aid comprehension.[1–2] Additionally, numerous hardware or software technologies, such as large buttons or eye tracking, can make documents accessible to those with motor impairments (see, for example, http://www.webaim.org/articles/motor/assistive.php#eyetracking).

The Digital Accessible Information System (Daisy) standard describes an open data format for the representation of interactive books that are accessible to those with print-related disabilities.[3] Daisy books may have both a textual and an audio component and allow for an active reading experience. To read a Daisy book, a reader needs a hardware or software playback system. Unlike a book on a cassette tape that users typically listen to from start to finish, readers using a Daisy book player can easily move backward and forward in the book; they can move to chapters, sections, pages, or bookmarks they have created. If the reader resumes reading, the player begins playback where the reader left off. Some players allow readers to change the magnification and the colors used. Many such players also highlight what is being spoken.

One of the problems that Daisy books are particularly good at solving is image translation. For example, in technical textbooks and journals, equations, diagrams, graphs, and other image-oriented material often accompany the text. These images are not immediately accessible to blind people. For those with limited vision, images can cause problems because they often don't look clear when magnified. And for those with cognitive disabilities, cluttered images or complicated ordering within the image might hinder understanding. Some images can be made accessible by providing a textual description of the image. However, for many scientific or mathematical images, this summary information doesn't help readers understand an image's structure and relationships. Daisy books can solve this problem by using structured markup instead of an image. For mathematical expressions, Mathematical Markup Language (MathML)[4] can represent the expressions; for diagrams and graphs, Scalable Vector Graphics (SVG)[5] can be used. These formats let users navigate the images, hear them as spoken words, or convert them to other formats, such as braille or tactile displays.

1070-986X/08/$25.00 © 2008 IEEE

Published by the IEEE Computer Society

While Daisy was designed to make materials accessible to those with print-related disabilities, the format is useful for people of all abilities. As such, the standard is an example of universal design; features such as being able to move to a section or a bookmark added by a reader are useful to everyone. The ability to have a visual, audio, or combined text-and-audio presentation generated from the same source material means that readers could choose an experience appropriate for their abilities or constraints. One example is using an audio-only version while driving a car or riding on mass transit. Because of Daisy's universal usability, several organizations have adopted the standard.

Evolution of Daisy

For the blind, braille translations of print books have provided an alternative for over a century. Braille books share many of the convenient features of print books, such as the ability to jump to pages or reread pages as desired. However, their weight, size, and binding make them hard to carry around. Braille books are large and heavy because the embossing paper is thicker than traditional paper; braille characters are approximately the same size as 36-point characters; and braille books must be loosely bound so as not to crush the braille dots. Reading a braille copy, of course, requires proficiency in braille. In some countries braille literacy has decreased significantly. For instance, US braille literacy declined significantly in the last 50 years from 50 percent of the school-aged blind children in the 1960s to 12 percent in the 1990s.[6]

In addition to the issues associated with braille, there are other factors that make a format other than traditional small-print books potentially useful to many people. For example, many older readers find it difficult to read regular print books due to normal aging processes and age-related eye diseases. Almost three million people in the US alone will suffer from macular degeneration by 2020.[7] Audio formats, such as talking books, provide an alternative to print books. In the 1970s, analog cassette tapes dominated the talking book industry. Cassette tapes and players were widely available and relatively inexpensive, and, unlike braille, don't require any special training to use. However, audio books lack the ability for readers to easily find sections or pages in a book. Being able to find a place in a book quickly is critical for studying or in a classroom setting where an instructor might say, ''Please turn to page 107.'' Also, the sharing of books between organizations or persons was hindered because of incompatible formats for the recordings used in different countries.

The Swedish Library of Talking Books and Braille (TPB) started the Daisy project in 1988. One of the chief goals was to move away from analog cassette tapes and toward a digital medium. In the early 1990s, compact discs gained widespread acceptance and TPB selected this format as the target medium for Daisy distribution. A prototype Daisy player was built in 1994 and presented at the International Conference on Computers Helping People with Special Needs (ICCHP) later that year. This resulted in considerable interest in the Daisy project.

In 1996, TPB invited other talking-book libraries to form a worldwide consortium to promote digital, talking books based on the Daisy ideas. In 1998, The consortium revised the format to conform to several World Wide Web Consortium (W3C) multimedia standards, such as XHTML and Synchronized Multimedia Integration Language (SMIL),[8] as the basis of a Daisy 2.0 specification. The following year some small enhancements led to new 2.01 and 2.02 specifications.

The original goals for Daisy centered on enhancing the audio experience for listening to e-books. The first Daisy players were audio-only; the majority of players remain audio-only today. However, it was recognized early on that Daisy could serve as a format for visual players too. Those players could offer extensive functionality, such as image display, text magnification, and synchronization of text and audio. People with limited vision and learning disabilities might find such features helpful.

The Daisy 2.02 specification requires many cross-references between the textual XHTML file and the controlling SMIL file. Producing this cross-referencing by hand for real books isn't feasible. Hence, use of the Daisy format required waiting for the arrival of tools that allowed for easier production of Daisy materials. The first such professional tool was announced in 2000. Production of Daisy material began a short while after that.

Feedback from users, producers, and Daisy player manufacturers, along with a push to make Daisy a formal standard, led to a revision of 2.02 becoming the ANSI/NISO Z39.86-2002 standard. This standard was updated in 2005 to become Z39.86-2005 and is commonly referred to as Daisy 3. The standard moved from using XHTML in Daisy 2 to using XML in Daisy 3, a switch that allowed for a much richer use of elements. Daisy 3 is a functional superset of Daisy 2.

An important feature of the Daisy 3 specification is that it allows for modular extensions. The first of these extensions was for mathematics and was approved in 2007.[9] The math extension is based on MathML, a W3C recommendation for encoding math in XML. The addition of a mathematics module to Daisy shows that Daisy can evolve by incorporating standards from other areas. Video and digital rights management are two such areas being considered. The Daisy Consortium closely monitors developments regarding XML and is continually gathering requirements for future versions.

Daisy not only is incorporating other standards, but also is being adopted by other organizations. The International Digital Publishing Forum's Open Publication Structure for e-books incorporates Daisy into its IDPF standard as its preferred vocabulary.[10] IDPF includes large publishers such as McGraw-Hill, John Wiley & Sons, and HarperCollins. The National Instructional Materials Accessibility Standard (better known as NIMAS) is based on Daisy and is the legally mandated format for new K–12 textbooks and supplementary material in the US.[11] Extension of these requirements to higher education textbooks likely will happen in the future.

Anatomy of a Daisy 3 book

A Daisy book can consist of audio only, text and images only, or a combination of both. All book types use the same set of file types, although some of the files are optional. Unless otherwise specified in our following discussion, all file types are based on XML. The nine file types comprising a Daisy 3 book are

- package file,
- textual content file,
- image file,
- audio file,
- synchronization file,
- navigation control file,
- resource file,
- presentation style file, and
- transform file.

The package file is a required type that acts as the master file of the Daisy book during playback. It contains metadata about the book such as the language, publisher, and copyright date; a manifest listing all the files comprising the book; and a <**spine**> element that indicates the book's linear reading order.

A textual content file contains the book's text and is referred to as the *DTBook file*. An audio-only book doesn't contain text and therefore will not include a DTBook file. The text content is marked up with a set of elements, similar to HTML, that indicate the information's structure and semantics. This markup is also known as *Daisy XML*. Daisy includes about 80 elements[3] such as

- <**sent**> and <**p**> to indicate sentences and paragraphs;
- <**level1**> through <**level6**> to indicate structural components, such as chapters and sections;
- <**img**> and <**imggroup**> to include images and image descriptions; and
- <**pagenum**>, <**noteref**>/<**note**>, and <**sidebar**> to mark page numbers, footnotes, and sidebars.

For the richest reading experience, the DTBook file should contain the full text, including footnotes and descriptions of all of the images in the book.

A Daisy book might optionally have images, stored in separate files. Daisy supports common image formats such as SVG, PNG, and JPEG. A Daisy book might contain recorded audio created by a person narrating or a synthetic speech voice. Daisy supports common audio formats such as MP3 and WAV.

When a Daisy book contains text, audio, and images, the synchronization file specifies the timing and order of presentation. The file conforms to SMIL 2.0 and is referred to as the *SMIL file* (pronounced ''smile''). The body of the SMIL file consists of a series of <**seq**> and <**par**> elements. Children of a <**seq**> element are presented sequentially to the reader and children of a <**par**> are presented simultaneously (in parallel). In a Daisy book the children may be

- <**text**> elements that reference content in the DTBook file,
- <**img**> elements that reference image files, or
- <**audio**> elements that reference content in the audio file.

For example, a <**par**> element might contain a <**text**> element that refers to a particular <**sent**> element in the DTBook file. The <**text**> element might be followed by an <**audio**> element that refers to a particular portion of an audio file by its starting and ending clock values. When the <**par**> element is played back, the player simultaneously displays the <**sent**> element of the DTBook file and plays the section of the audio file referenced in the <**audio**> element. If the <**text**> and <**audio**> elements are children of a <**seq**> instead of a <**par**>, then the <**sent**> element is displayed first, followed by playback of the audio file. Image playback is handled similarly by using the <**img**> element. Figure 1a (next page) shows an example of a SMIL file controlling text and audio files. The <**seq**> and <**par**> elements can be nested in other <**seq**> and <**par**> elements to provide any required combination of sequential and parallel presentation of text, images, or audio.

The navigation control file, referred to as the *NCX file*, supports navigating to particular parts of the book. The NCX file summarizes the significant structures of the book. One part of the NCX information acts as a hierarchical table of contents. Another part of the file links to the book's pages. This functionality permits the reader to browse the page list or jump directly to a given page. The NCX file might contain lists of illustrations, tables, equations, or other items of interest. The reader can browse a given list and jump directly to an item. Items in a list don't need to be of the same type; a list might be named *pendulum experiment*, for example, and refer to associated text, equations, figures, and tables.

The resource file specifies text, audio clips, or images for notifying the reader when encountering certain book features or structures. For example, the resource file makes it possible to have page numbers audibly announced if the user chooses to hear them.

In addition to these features, Daisy books support formatting through both cascading stylesheets (CSS) and Extensible Style Sheet Language (XSL) styling, as noted in the presentation style file. The styling can be applied to the visual, audio, or braille presentation of the book's content. Transform files are used when some information in a Daisy book includes a simplified fallback format for players that aren't able to handle advanced content. Mathematics in the form of MathML, for example, has an image and text fallback format. The Daisy book can include XSLT to perform transformations into a simplified format.

The Daisy standard describes two other file types not included in the previous list and not part of the book itself. A media unit, such as a CD-ROM or flash drive, might contain more than one Daisy book. Conversely, a large book might reside on several media units. In these situations, the distribution information file describes how to locate the necessary information for each book. And for cases where readers want to set bookmarks at points of interest, they can do so with their own text or audio notes, which are referenced in a bookmark file and are sharable with other people.

For reasons of performance or convenience, each type of information can be divided among several files. For example, the content of a large book may be split into several DTBook files. Also, when a book spans several CDs, there are further constraints on the files to minimize swapping between the CDs. Some information, like the package file, is duplicated on each CD.

Daisy book production

Typically, Daisy book production begins with an existing print book. There are different types of production processes: starting with text or starting with audio.