Accessibility of Documents

Georgios Kouroupetroglou
National and Kapodistrian University of Athens, Greece

INTRODUCTION

Printed or electronic documents include mainly books, newspapers, journals, periodicals, webpages, reports, articles, letters, pamphlets, and e-mails. During the past four decades, the evolution of computer science and information technology brought out new perspectives for documents. Nowadays, the vast amount of documents is edited and stored in an electronic or digital format. Moreover, printed documents of the past can be digitized using common scanners and OCR (optical character recognition) software.

Document accessibility enables an electronic document, either in the form of a common file format (such as .pdf .doc, .ppt) or a Web document, to be used effectively, efficiently and satisfactory by more people in more situations or context of use. It concerns all the aspects of document functionality that includes browsing, searching, navigation and reading (Doermann, Rivlin, & Rosenfeld, 1998).

Accessibility of documents is very important not only for the print-disabled readers, i.e. those with vision impairment (blindness, low vision, color blindness, dyschromatopsia, etc.), a learning disability (including dyslexia) or a motor disability (such as loss of dexterity that prevents the physical handling of a document), but also for those with an occasional or situational "disability." A typical example is the car driver whose eyes and hands are busy.

In this article, after the analysis of a document's architecture, we present the main principles and technological topics of document accessibility. Finally, the emerging area of Document-to-Audio (DtA) is presented. DtA essentially constitutes the next generation of Text-to-Speech systems and they are capable to provide much better accessibility to documents.

BACKGROUND

According to the communication theorist Marshall McLuhan, a document is the "medium" in which a "message" (information) is communicated (McLuhan & Fiore, 2005). The content of the printed or electronic documents includes mainly the text and the images (i.e. figures, drawings, graphs, pictures, charts, diagrams, maps, photos, etc.). Furthermore it may include mathematical or in general scientific symbols and formulas. The term *text-document* refers to the textual content only of a document.

Besides its content, a printed or electronic document contains a number of *presentation elements* or *attributes* that apply on its text content: a) design glyphs or typographic elements (i.e. visual representation of letters and characters in a specific font and style) and b) arrangement of the content on the page and the document as a whole. For example, the title of a chapter can be recognized by placing it at the top of the page and in larger font size than the body of the text. Also, text color, but also the bold font style, can be used to indicate emphasis in a specific part of a text-document.

The elements of a text-document can be classified in three layers (Kouroupetroglou & Tsonos, 2008):

- **Logical Layer:** It associates content with structural elements such as headings, titles/subtitles, chapters, paragraphs, tables, lists, footnotes, and appendices.
- **Layout Layer:** It associates content with architectural elements relating to the arrangement on pages and areas within pages, such as margins, columns, alignment and orientation (portrait or landscape).
- **Typography Layer:** It includes font (type, size, color, background color, etc.) and font style such as bold, italics, underline. In contrast to the *rich text*, the term *plain text* indicates text

of in any unique font type and size, but without font style.

The above three layers are complementary and not independent. Typography can be applied to both the logical and the layout layers of a document. Moreover, typography can be applied to the main body of the text directly. For example, a word in bold can be used either for the introduction of a new term or to indicate a person's name. Also, a heading can be arranged in the center of a line (layout layer).

The organization of a document can be classified into two main aspects: the logical and the physical. The logical layer of the document architecture defined above corresponds to its logical organization with the same elements (e.g. headings, titles/ subtitles, chapters, paragraphs, tables, lists, footnotes, and appendices). At the page level, the physical organization of a document is described by its layout layer in connection with the physical realization of a number of logical layout elements (e.g. headings, titles/ subtitles, paragraphs, tables, lists, footnotes). The organization of a printed of electronic multipage document as a whole corresponds with the physical implementation of a part of its logical layer elements (e.g. chapters, appendices, indexed, references). The organization of a document is domain specific (e.g. text book, scientific paper, technical report, newspaper, magazine).

The term *text signal* has been proposed (Lorch, 1989) as the writing device that emphasizes aspects of a text's content or structure carrying semantic information over and above the content. It attempts to pre-announce or emphasize content and/or reveal content relationship (Lemarié, Eyrolle, & Cellier, 2006). Headings or titles in text-documents are considered as signals (Lorch, Chen, & Lemarié, 2012).

DOCUMENT ACCESSIBILITY

The accessibility of documents can be analyzed in the following four dimensions: technological, legal, economic and social/ethical.

The Convention on the Rights of Persons with Disabilities (CRPD, 2012) of the United Nations, adopted on 13 Dec. 2006 by the General Assembly and effective from 3 May 2008, and specifically Article 9, is the main international legal instrument for document accessibility. Up the end of 2012, 126 countries ratified and adopted CRPD on a national legislation level. The Americans with Disabilities Act (ADA, 2012) enacted by the U.S. Congress in 1990 as well as the Federal procurement of electronic and information technology (Section 508, 2012) cover the issue of document accessibility. i2010 (2012), the European Union's respective policy framework for the information society and media 2005-2009, along with the Digital Agenda for Europe (Europe 2020, 2012), incorporate the Commission's strategy involving with document accessibility.

The economic aspects of document accessibility include on one hand the development cost of accessible documents and on the other the return from their production and distribution (e.g. increase the range of possible customers and/or decrease the production of alternative formats). By applying the universal design principle, the extra cost one has to spend for a document to become accessible can be sufficiently small. *Universal access* embraces theoretical, methodological, and empirical research of both a technological and non-technological nature that addresses accessibility, usability, and, ultimately, acceptability of information society technologies by anyone, anywhere, at any time, and through any media and device (Stephanidis, 2001). In the framework of universal access (Stephanidis et al., 1988), we define *universal design* or *design-for-all* as the general framework catering for conscious and systematic efforts to proactively apply principles, methods, and tools to develop documents that are accessible and usable by more people, including the youngest, the elderly and people with different types of disabilities, in more situations or context of use, thus avoiding the need for a posteriori adaptations, or specialized design.

Social/ethical issues of document accessibility promote equal opportunities, without exclusion and are part of the Corporate Social Responsibility of an institution, either public or private.

The basic technological requirements for a document to be accessible are:

- Content must be *perceivable*. Information must be presentable to users in ways (such as other modalities) they can perceive, i.e. content cannot be unobservable to all of their senses.
- The reading platform/agent (hardware and/ or software) must be *operable* concerning

Related Content

Revolution in Military Affairs as Technology-Driven Change in Military Operation
Rafa Kope (2015). *Encyclopedia of Information Science and Technology, Third Edition (pp. 6534-6542).*
www.irma-international.org/chapter/revolution-in-military-affairs-as-technology-driven-change-in-military-operation/113112

Construction of Building an Energy Saving Optimization Model Based on Genetic Algorithm
Xin Xuand Xiaolong Li (2023). *International Journal of Information Technologies and Systems Approach (pp. 1-15).*
www.irma-international.org/article/construction-of-building-an-energy-saving-optimization-model-based-on-genetic-algorithm/328758

An Eco-System Architectural Model for Delivering Educational Services to Children With Learning Problems in Basic Mathematics
Miguel Angel Ortiz Esparza, Jaime Muñoz Arteaga, José Eder Guzman Mendoza, Juana Canul-Reichand Julien Broisin (2019). *International Journal of Information Technologies and Systems Approach (pp. 61-81).*
www.irma-international.org/article/an-eco-system-architectural-model-for-delivering-educational-services-to-children-with-learning-problems-in-basic-mathematics/230305

Sampling, Channels, and Contact Strategies in Internet Survey
Ester Macrìand Cristiano Tessitore (2013). *Advancing Research Methods with New Technologies (pp. 34-47).*
www.irma-international.org/chapter/sampling-channels-contact-strategies-internet/75938

PRESCAN Adaptive Vehicle Image Real-Time Stitching Algorithm Based on Improved SIFT
Qian Li, Yanli Xuand Pengren Ding (2023). *International Journal of Information Technologies and Systems Approach (pp. 1-17).*
www.irma-international.org/article/prescan-adaptive-vehicle-image-real-time-stitching-algorithm-based-on-improved-sift/321754