

Monika Fiedorowicz, Agnieszka Janucik

## A different approach to accessibility, i.e. WCAG in full-text documents – working with files

O otwartości inaczej, czyli WCAG w dokumentach pełnotekstowych – praca z plikami

Main Library, Medical University of Białystok, Poland  
Correspondence to: monika.fiedorowicz@umb.edu.pl

**SUMMARY:** Polish Platform of Medical Research provides open access to scientific resources, including full text publications, PhD theses and other repository documents. One of the assumptions behind the project was to adapt all files posted on the platform to the WCAG (Web Content Accessibility Guidelines) to reach a wide audience, including people with disabilities. To implement that assumption, we had to restructure all files, describe their headings, and meet several other requirements laid down in the WCAG. Each file had to be adjusted separately before uploading it to the platform, which was often challenging and time-consuming. The article outlines the principles for tagging PDF files to convert them in line with the WCAG.

**STRESZCZENIE:** Polska Platforma Medyczna zapewnia otwarty dostęp do zasobów nauki, w tym do pełnych tekstów publikacji, rozpraw doktorskich i innych dokumentów zawartych w repozytorium. Jednym z założeń projektu było to, by wszystkie pliki zamieszczone na platformie zostały dostosowane do standardów WCAG (Web Content Accessibility Guidelines) w celu dotarcia do szerokiego grona odbiorców – także do osób niepełnosprawnych. Aby zrealizować to założenie, należało nadać plikom odpowiednią strukturę, opisać nagłówki i spełnić szereg innych wytycznych ujętych w dokumencie WCAG. Przygotowanie danych do zdeponowania wymagało oddzielnego dostosowania każdego pliku, co często było pracą trudną i bardzo czasochłonną. W artykule zaprezentowano zasady znakowania plików PDF w celu konwersji do formatu WCAG.

**KEYWORDS:** Polish Platform of Medical Research, WCAG, web accessibility, aids for disabled

*This is an open access article under the CC BY-NC-ND license.*

Polish Platform of Medical Research (PPM) is a portal that presents and promotes scientific achievements and the potential of its partners in medicine, pharmacy, dentistry, public health, safety, occupational hygiene, ergonomics, and health protection. One of the objectives was to increase the accessibility of scientific resources for people with disabilities. Therefore, all full-text files published on the platform have been adapted to the WCAG.

People who are blind, visually impaired, deaf and hearing impaired, deaf-blind, with intellectual or motor disabilities, or having problems with their arms or hands which make it difficult to use the mouse, are at particular risk of digital exclusion. Most of them use the keyboard, keyboard shortcuts and text-to-speech software when interacting with computers and websites. The inaccessibility of keyboard navigation poses a major problem. Some visually impaired people need special software to enlarge screen content, while for others an adequate contrast ratio is enough. A document can be deemed accessible if people with disabilities may access the full content and navigate or interact with the site comfortably using the keyboard. A website and documents posted on it must have a certain structure along with

headings described, and must follow many other guidelines set out in the Web Content Accessibility Guidelines (WCAG).

The WCAG specification contains requirements for the content of websites for people with disabilities. Web accessibility was first discussed at the 1994 World Wide Web conference in Chicago. Guidelines were developed a year later at the University of Wisconsin-Madison's Trace Centre. Once standardised, those guidelines became a source material for the WCAG. Work on the WCAG principles began in 1996 at the initiative of the Web Accessibility Initiative (WAI) of the World Wide Web Consortium (W3C). The goal was to develop solutions for developers and web users. WCAG 1.0 were published in 1999. Their main principle was to make content accessible to as many people as possible. The dynamic growth of the Internet forced the adaptation of WCAG 1.0 to the new reality. A new version of the guidelines, WCAG 2.0, with more precise requirements for the accessibility of websites for users, including users with disabilities, was published in 2008. It became the international ISO/IEC 40500:2012 standard in 2012. Two years later, some guidelines were incorporated in the European standard EN 301 549. In 2018, the

standard was updated to 2.1 and the guidelines were adapted to all new technical solutions. The need to implement WCAG 2.1 was justified by the European Parliament and Council in Directive 2016/2102 of 26 October 2016, which translated into amendment in Polish law. The Act of 4 April 2019 on the digital accessibility of websites and mobile applications (Dz.U. [Polish Journal of Laws] of 2019, item 848) sets forth the procedures to be implemented in the event of evading obligations, and contains information on the accessibility statements and guidelines that include new technological solutions such as mobile applications.

The WCAG standard is based on four principles: websites must be perceivable, operable, understandable, and robust. **Perceivable** means that information and interface components must be presentable to users in ways they can perceive. We use sight and hearing to perceive information on the Internet, so this requirement mainly applies to people with impairments of these senses. **Operability** is the second principle. User interface components and navigation must be operable and available from a keyboard. According to the third principle, the web content and operation of the user interface must be **understandable**. And finally, the content must be **robust** so that it can be interpreted correctly by various users, taking into account the capabilities of assistive technologies.

In order to obtain a WCAG-compliant file, we must pay attention to the following elements: document title; document language; document structure (headers, sections); alternate text added to non-text elements – images, graphics, charts; tables; lists of items; contrast ratio between text and background and between text and graphics.

Document title is the first element processed by the text-to-speech software. Titles should be placed in metadata which, depending on the software used, are entered in a special document information window. When the title is missing, assistive technologies usually read the filename, which often means nothing to a person with disability.

Document language is another important element. It must be specified so that the text can be read correctly by the assistive technologies and appropriate dictionaries are used for spell checking.

Headers are the key components that make a document accessible. In the files we prepared, all section and chapter titles had to be added as headers of different levels, with at least one header 1 (H1) in each document. This is how the correct structure was created; with that structure blind users who rely on text-to-speech software are able to move more easily between headers and very quickly get to know the content of the document without having to read it all. When working with a document, they use keyboard shortcuts (H – header, T – table,

G – graphics, etc.). If a user presses the H key several times, he/she will move to the next headers. The figure below shows an example of the document structure.

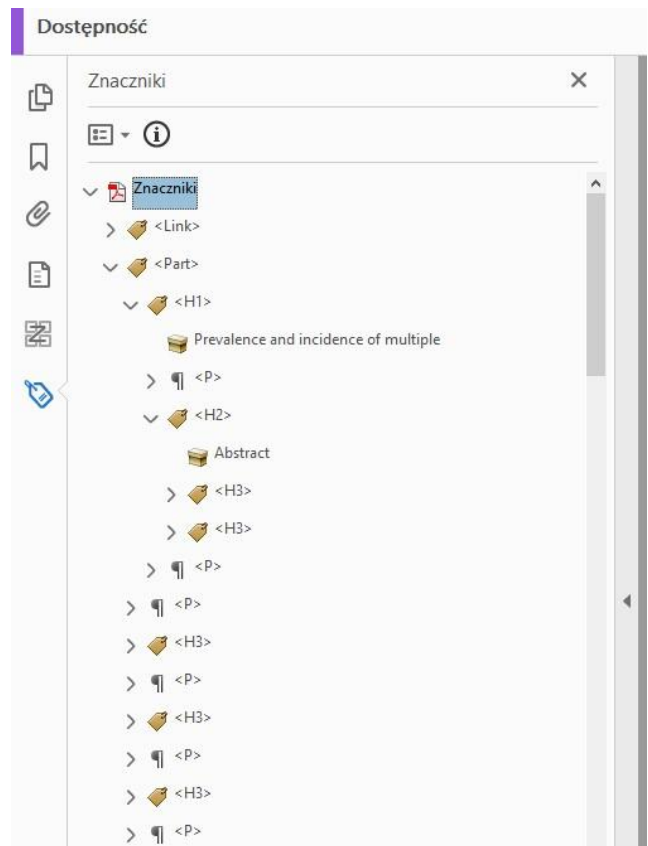


Fig. 1. Tag tree in Adobe Acrobat Pro

Text documents often contain graphics: photographs, figures, diagrams and charts. If a graphic contains important information, it should be provided with alternate text, which must be worded accordingly and provide a person with a visual impairment with a complete set of information. Alternate text is not visible to the reader and users with disabilities will get to know it by using the text-to-speech software. Therefore, an accurate description of the content is very important. Purely decorative graphic elements that do not carry any information and can be labelled as artefacts are exceptions to the alternate text principle. Assistive software, such as screen readers or speech synthesizers, cannot see the graphics but are able to read the text information contained in the “alternate text” field. An example of placeholder text is shown in fig. 2.

Tables are another important element. Each table must have a title, while columns and rows should be tagged as headers. Tables should be created using built-in tools instead of tabs, spaces or line breaks. All tables in the prepared documents had to be checked for correct markup of header cells, rows, data cells and regularity of the table. In the event of discrepancies, errors had to be corrected manually.

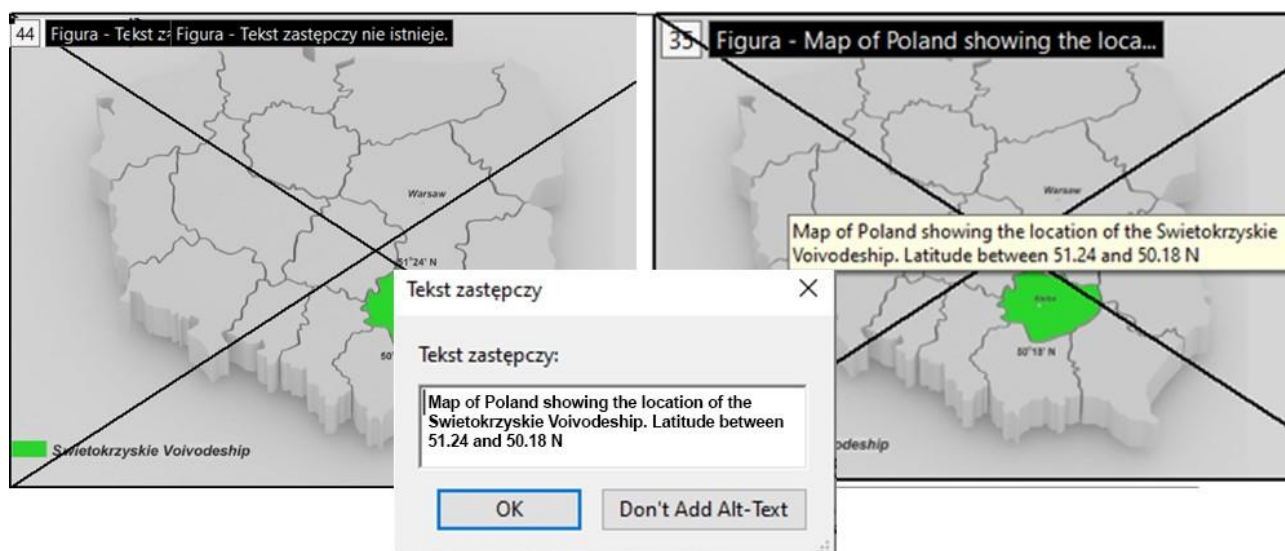


Fig. 2. Adding alternate text to an image in Adobe Acrobat Pro

Lists of items are an important factor that contributes to the better orientation of the user on the website intended for people with disabilities. They are easily readable for screen readers, allowing the blind to freely move between the lists. A list must be structured accordingly to guarantee this functionality. A list (L) consists of list items (LI), while list items contain labels (Lb1) and a body (Lbody), i.e. a description of a given sub-item. If the text we worked on contained a manually created list, without a built-in functionality, we had to change it using tags that helped control the structure and levels of the list. This made it possible to change the order or types of the list items and to create new items. The list had the correct structure and retained the required block level items.

A vast majority of people with visual impairment use screen magnification software. Therefore, according to the WCAG, the contrast ratio between text (and images of text) and background should amount to 4.5:1. A simple software (such as Contrast Analyser) can be used to check the contrast ratio. After selecting two adjacent colours, the contrast ratio is shown and the user can check if the condition is met. When the “colour contrast” principle fails, document content should be changed by using colours with a high contrast ratio.

When preparing text document files for uploading into the Polish Platform of Medical Research, we used two programs: Microsoft Word and – in a vast majority of cases – Adobe Acrobat Pro. Adobe Acrobat offers tools that support the adaptation of files to the WCAG. They allow users to verify and enhance accessibility in existing PDF files. Each document must have a tag structure with a certain hierarchy that determines reading order. Tags are added to headers, paragraphs, tables, pictures, lists of items, and other document components. When there are no tags, Adobe Acrobat Pro allows the user to automatically tag a given document. The automatic check of PDF accessibility proved a useful fea-

ture when adapting files to the WCAG standard. The tool runs a preliminary check of whether a given document meets the accessibility conditions. It produces a report with the errors described. When preparing the files for PPM, we had to review all comments and remove any defects that required manual correction. It was necessary to manually check the reading order and correct tagging of headers and paragraphs. Each non-text item, such as an image or graph, had to be provided with an alternate text or tagged as an artefact. Tables required a lot of effort, as it was necessary to manually tag header rows and columns and data cells.

Each file adapted in line with the WCAG required an individual approach. We have encountered numerous issues which resulted from the complexity of documents or various ways in which text layout was prepared. Sometimes it was necessary to interfere with document appearance in order to improve the contrast ratio. The teams formed by the project partners had to prepare a total of over 20,000 files. All files deposited on the PPM are in the Repository. The files adapted to WCAG 2.1 are marked with the accessibility icon.



Polska Platforma Medyczna (PPM) to portal prezentujący i promujący osiągnięcia naukowe oraz potencjał partnerów w zakresie medycyny, farmacji, stomatologii, zdrowia publicznego, bezpieczeństwa, higieny pracy, ergonomii i ochrony zdrowia. Jednym z założeń było zwiększenie dostępności do zasobów nauki dla osób niepełnosprawnych, dlatego wszystkie pliki dokumentów pełnotekstowych publikowane na platformie zostały dostosowane do standardu WCAG.

Sz szczególnie narażone na wykluczenie cyfrowe są osoby niewidome, niedowidzące, głuche i słabosłyszące, głuchoniewidome, z niepełnosprawnością intelektualną lub ruchową, mające pro-

blemy z rękoma czy dłońmi, co utrudnia obsługę myszki. Większość takich osób w interakcji z komputerem i serwisami internetowymi posługuje się głównie klawiaturą, korzysta ze skrótów klawiaturowych i programów czytających. Niedostępność nawigacji z poziomu klawiatury jest zatem dużym problemem. Część osób niedowidzących potrzebuje specjalnych programów powiększających zawartość ekranu, innym natomiast wystarczy odpowiedni kontrast. Dokument nazywamy dostępnym wtedy, gdy osoby niepełnosprawne mają pełny dostęp do treści i mogą skorzystać z wygodnej nawigacji czy interakcji z serwisem za pomocą klawiatury. Strona internetowa i zamieszczane na niej dokumenty muszą mieć odpowiednią budowę i opisane nagłówki oraz spełniać wiele innych wytycznych zawartych w dokumencie WCAG (Web Content Accessibility Guidelines).

Specyfikacja WCAG zawiera wymagania dotyczące treści serwisów internetowych dla osób niepełnosprawnych. O dostępności sieci mówiono po raz pierwszy na konferencji World Wide Web w Chicago w 1994 r. Rok później na Uniwersytecie Wisconsin-Madison, w centrum Trace, powstały wytyczne, które po ujednoczeniu stały się materiałem źródłowym dla WCAG. Prace nad zasadami WCAG rozpoczęły się w 1996 r. z inicjatywy zespołu Web Accessibility Initiative (WAI), pracującego w World Wide Web Consortium (W3C). Efektem prac miały być rozwiązania dla programistów i użytkowników sieci. W roku 1999 powstał standard WCAG 1.0, którego główną zasadą było udostępnianie treści jak największej liczbie odbiorców. Dynamiczny rozwój Internetu wymusił dostosowanie wymogów WCAG 1.0 do nowych realiów. W roku 2008 wprowadzono nową wersję standardu, WCAG 2.0, zawierającą bardziej precyzyjne wymagania co do dostępności witryn dla użytkowników – również dla osób z niepełnosprawnością. W 2012 r. stała się ona międzynarodową normą ISO/IEC 40500:2012. Dwa lata później część wytycznych zawarto w normie europejskiej EN 301 549. Rok 2018 przyniósł aktualizację standardu do wersji 2.1, wytyczne zostały dopasowane do wszystkich nowych rozwiązań technicznych. Zasadność wdrażania standardu WCAG 2.1 potwierdziły Parlament Europejski i Rada Unii Europejskiej w dyrektywie 2016/2102 z 26 października 2016 r., co przełożyło się na nowelizację polskiego prawa. Ustawa z dnia 4 kwietnia 2019 r. o dostępności cyfrowej stron internetowych i aplikacji mobilnych podmiotów publicznych (Dz.U. 2019 poz. 848) określa procedury wdrażane w przypadku uchylania się od obowiązków, zawiera informacje o deklaracji dostępności i wytyczne uwzględniające nowe rozwiązania technologiczne, takie jak aplikacje mobilne.

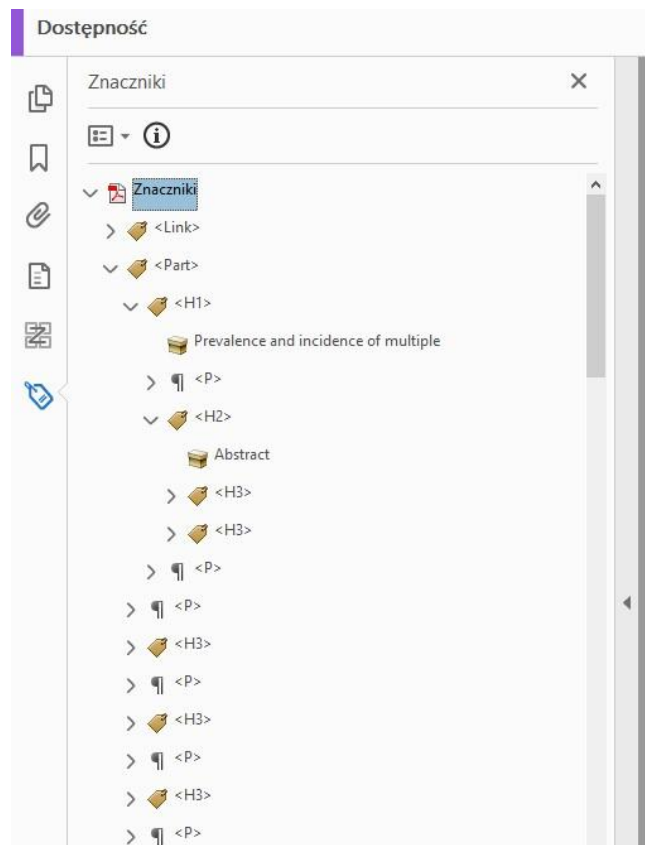
Standard WCAG opiera się na czterech zasadach: postrzegalności, funkcjonalności, zrozumiałości i solidności. **Postrzegalność** oznacza, że informacje i komponenty interfejsu muszą być przedstawione w sposób dostępny dla zmysłów użytkownika. Do odbioru informacji w Internecie używamy wzroku i słuchu, powyższy wymóg odnosi się więc głównie do osób z zaburzeniami tych

zmysłów. Druga zasada to **funkcjonalność** – komponenty interfejsu użytkownika i nawigacja muszą pozwalać na interakcję i być dostępne z poziomu klawiatury. Zgodnie z zasadą **zrozumiałości** treści zamieszczona w serwisie i obsługa interfejsu użytkownika muszą być zrozumiałe. Wreszcie zasada **solidności** wymaga rzetelności treści, tak aby mogła ona być poprawnie interpretowana przez wielu różnych użytkowników, biorąc pod uwagę możliwości technologii asystujących.

Aby uzyskać plik zgodny z zasadami WCAG, musimy zwrócić uwagę na następujące elementy: tytuł dokumentu; język dokumentu; strukturę dokumentu (nagłówki, sekcje); tekst alternatywny dodany do elementów nietekstowych – obrazków, grafik, wykresów; tabele; listy elementów; kontrast między tekstem a tłem i między tekstem a grafikami.

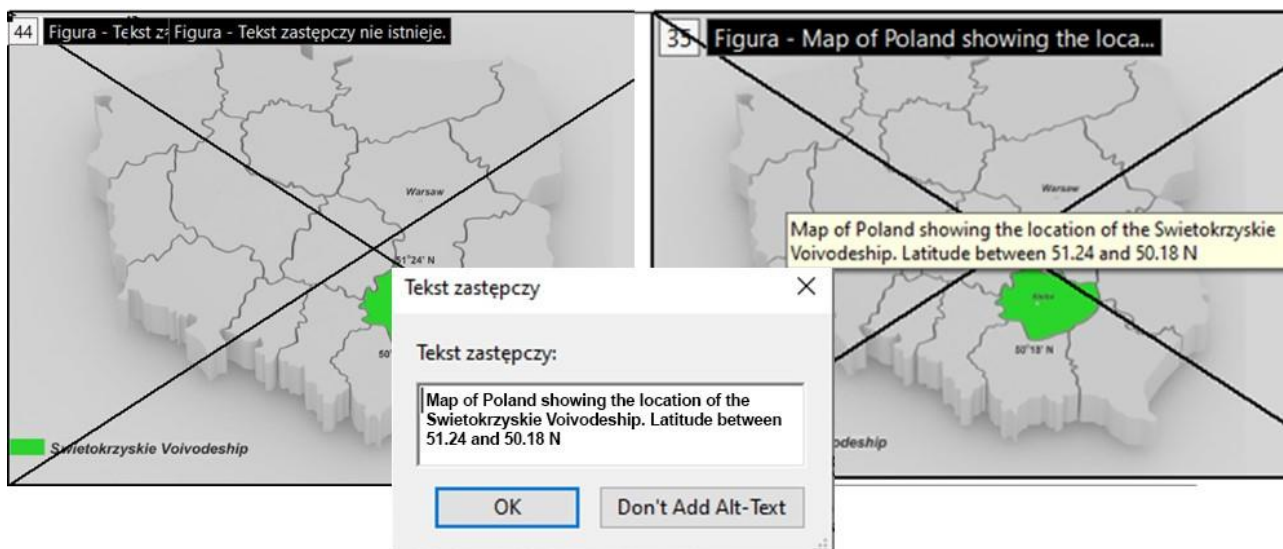
Pierwszym elementem przetwarzanym przez programy czytające jest tytuł dokumentu. Tytuły powinny być umieszczane w metadanych, które w zależności od używanego programu wypełnia się w specjalnym oknie informacji o dokumencie. Gdy brakuje tytułu, technologie pomocnicze czytają zwykle nazwę pliku, która często nic nie mówi osobie niepełnosprawnej.

Kolejny ważny element to język dokumentu, który trzeba wybrać tak, aby tekst został poprawnie przeczytany przez technologie wspomagające, a przy sprawdzaniu pisowni były używane odpowiednie słowniki.



Il. 1. Drzewo znaczników w programie Adobe Acrobat Pro





Il. 2. Dodawanie tekstu alternatywnego do obrazka w programie Adobe Acrobat Pro

Kluczowym komponentem świadczącym o dostępności dokumentu są nagłówki. W przygotowywanych plikach wszystkie tytuły sekcji i rozdziałów należało oznaczyć nagłówkami różnych poziomów, w każdym dokumencie musiał się znaleźć przynajmniej jeden nagłówek 1 (H1). Tak powstawała poprawna struktura, dzięki której osoby niewidome, korzystające z programów czytających, mają możliwość łatwiejszego przemieszczania się między nagłówkami i bardzo szybko poznają treść dokumentu, bez potrzeby czytania całości. Podczas pracy z dokumentem używają skrótów na klawiaturze (H – nagłówek, T – tabela, G – grafiki itp.). Kilukrotne przyciśnięcie klawisza H skutkuje przejściem do kolejnych nagłówków. Na il. 1 przedstawiono przykładową strukturę dokumentu.

W dokumentach tekstowych często zamieszczane są grafiki: zdjęcia, ryciny, wykresy, diagramy. Jeżeli grafika zawiera istotne informacje, powinna zostać opatrzona opisem alternatywnym, który musi być odpowiednio sformułowany i zapewniać osobie z niepełnosprawnością wzroku komplet informacji. Opis alternatywny nie jest widoczny dla czytelnika, a użytkownicy z niepełnosprawnościami poznają go za pomocą programów czytających, dlatego wiarygodny opis zawartości ma duże znaczenie. Odstępstwem od zasady umieszczania tekstu zastępczego są elementy graficzne wyłącznie dekoracyjne, które nie niosą żadnej informacji i mogą być oznaczane jako artefakty. Oprogramowanie wspomagające – czytniki ekranowe czy syntezatory mowy – nie widzą grafiki, mogą jednak odczytać informację tekstową zawartą w polu „tekst zastępczy”. Przykład tekstu zastępczego znajduje się na il. 2.

Kolejnym ważnym elementem są tabele. Każda musi mieć wprowadzony tytuł, a kolumny i wiersze muszą zostać opisane nagłówkami. Tabele powinny być tworzone za pomocą wbudowanych narzędzi, a nie przy wykorzystaniu tabulatorów, spacji czy znaków końca linii. Wszystkie zestawienia tabelaryczne w przygotowywanych dokumentach wymagały sprawdzenia popraw-

ności identyfikacji komórek nagłówkowych, rzędów i komórek danych oraz regularności tabeli. W przypadku niezgodności należało ręcznie poprawić błędy.

Istotnym czynnikiem, który sprzyja lepszej orientacji użytkownika w serwisie dla osób niepełnosprawnych, są listy elementów – bardzo dobrze rozpoznawane przez czytniki ekranu, dzięki czemu niewidomi mogą swobodnie poruszać się między listami. Zachowanie funkcjonalności gwarantuje odpowiednia budowa listy. Lista (L) składa się z elementów listy (LI), a elementy listy zawierają etykiety (Lbi) i treść listy (Lbody), czyli opis danego podpunktu. Jeżeli w opracowywanym tekście znajdowała się lista utworzona ręcznie, bez użycia wbudowanej opcji, trzeba było to zmienić za pomocą znaczników, które pomagały w kontrolowaniu struktury i poziomów listy. Pozwalało to na zmianę kolejności lub typów elementów listy oraz tworzenie nowych elementów. Dzięki temu lista ma poprawną strukturę i zachowuje wymagane elementy poziomu blokowego.

Zdecydowana większość osób niedowidzących posługuje się programami powiększającymi – dlatego w standardzie WCAG istotny jest kontrast między tekstem lub grafikami tekstowymi a tłem, który powinien wynosić 4,5 : 1. Do sprawdzenia kontrastu można użyć prostego programu (np. Contrast Analyser). Po zaznaczeniu dwóch sąsiadujących kolorów otrzymujemy współczynnik kontrastu i możemy sprawdzić, czy warunek jest spełniony. Gdy reguła „kontrast koloru” zwraca błąd, należy zmienić zawartość dokumentu przez zastosowanie kolorów o wysokim kontraście.

Przygotowując pliki dokumentów tekstowych do zamieszczenia na Polskiej Platformie Medycznej, korzystaliśmy z dwóch programów: Microsoft Word oraz – w zdecydowanej większości przypadków – Adobe Acrobat Pro. Program Adobe Acrobat oferuje narzędzia wspomagające przystosowywanie plików do standardu WCAG. Umożliwiają one weryfikowanie i zwiększanie dostępności w istniejących plikach PDF. Każdy dokument musi mieć od-

powiednią strukturę znaczników, które tworzą pewną hierarchię, określającą kolejność czytania. Znacznikami opatruje się nagłówki, akapity, tabele, obrazki, listy elementów i inne składowe dokumentu. Gdy znaczników nie ma, Adobe Acrobat Pro umożliwia automatyczne oznakowanie dokumentu. Funkcją przydatną podczas przekształcania plików do standardu WCAG było także automatyczne sprawdzanie dostępności pliku PDF. Narzędzie to wstępnie ocenia, czy dokument spełnia warunki dostępności – wynikiem jest raport z opisanymi błędami. W trakcie przygotowywania plików na potrzeby PPM należało przejrzeć wszystkie uwagi i usunąć usterki wymagające ręcznej korekty. Ręczne sprawdzanie było konieczne w przypadku kolejności odczytu dokumentu oraz poprawności oznakowania nagłówków i akapitów. Każdy element nietekstowy – obrazek lub wykres – musieliśmy opatrzyć tekstem alternatywnym bądź oznaczyć jako artefakt. Dużego nakładu pracy wymagała praca z tabelami, często bardzo rozbudowanymi, w których należało ręcznie oznaczyć wiersze i kolumny nagłówkowe oraz komórki danych.

Każdy plik dostosowywany do standardu WCAG wymagał indywidualnego podejścia. Napotykalismy wiele problemów, które wynikały z różnorodnych sposobów składania tekstu czy złożoności dokumentów. Czasami niezbędna okazywała się ingerencja w wygląd dokumentu, służąca poprawie kontrastu. Zespoły

utworzone przez partnerów projektu musiały przygotować łącznie ponad 20 tys. plików. Wszystkie pliki zdeponowane na Polskiej Platformie Medycznej znajdują się w Repozytorium. Pliki dostosowane do standardu WCAG 2.1 oznaczono ikoną dostępności.

## BIBLIOGRAPHY

1. Adobe Help Center [Internet]. Adobe; [cited 12.05.2021.] Available from: <https://helpx.adobe.com/>
2. Czym jest WCAG? [Internet]. Blog Proformat; [cited 12.05.2021.] Available from: <https://www.proformat.pl/blog/czym-jest-wcag/>
3. Intro to Web Accessibility Standards [Internet]. SBCTC's Library of Accessibility Resources; [cited 12.05.2021.] Available from: [https://sbctc.instructure.com/courses/1578604/pages/intro-to-web-accessibility-standards?module\\_item\\_id=26673454](https://sbctc.instructure.com/courses/1578604/pages/intro-to-web-accessibility-standards?module_item_id=26673454)
4. Marcinkowski A, Marcinkowski P. WCAG 2.0 Podręcznik Dobrych Praktyk: Fundacja Widzialni; 2012.
5. Paszkiewicz D, Dębski J. Dostępność serwisów internetowych : dobre praktyki w projektowaniu serwisów internetowych dostępnych dla osób z różnymi rodzajami niepełnosprawności. Warszawa: Stowarzyszenie Przyjaciół Integracji; 2013.
6. WAI early days [Internet]. Web Accessibility Initiative; [cited 12.05.2021.] Available from: <https://www.w3.org/WAI/history/>

---

### MONIKA FIEDOROWICZ

Main Library, Medical University of Białystok, Poland  
Correspondence: [monika.fiedorowicz@umb.edu.pl](mailto:monika.fiedorowicz@umb.edu.pl)

### AGNIESZKA JANUCIK

Main Library, Medical University of Białystok, Poland  
Correspondence: [agnieszka.janucik@umb.edu.pl](mailto:agnieszka.janucik@umb.edu.pl)