Analyse von manuellen Prüfmethoden zur Bewertung der Barrierefreiheit von Websites nach BITV 2.0

Masterarbeit im Studiengang
Computer Science and Media

vorgelegt von
Gisela Andrea Kollotzek
an der Hochschule der Medien Stuttgart
am 06. Mai 2021
zur Erlangung des akademischen Grades eines Master of Sicence

Erst-Prüfer:
Prof. Dr. Gottfried Zimmermann

Zweit-Prüfer: Tobias Ableitner, M. Sc.
Praxisbetreuerin: Anne-Marie Nebe

Die Arbeit ist in Kooperation mit der Firma T-Systems Multimedia Solutions GmbH entstanden.

Stuttgart, den 06. Mai 2021

Ehrenwörtliche Erklärung

Hiermit versichere ich, Gisela Andrea Kollotzek, ehrenwörtlich, dass ich die vorliegende Masterarbeit mit dem Titel: „Analyse von manuellen Prüfmethoden zur Bewertung der Barrierefreiheit von Websites nach BITV 2.0“ selbstständig und ohne fremde Hilfe verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen der Arbeit, die dem Wortlaut oder dem Sinn nach anderen Werken entnommen wurden, sind in jedem Fall unter Angabe der Quelle kenntlich gemacht. Die Arbeit ist noch nicht veröffentlicht oder in anderer Form als Prüfungsleistung vorgelegt worden.

Ich habe die Bedeutung der ehrenwörtlichen Versicherung und die prüfungsrechtlichen Folgen (§ 26 Abs. 2 Bachelor-SPO (6 Semester), § 24 Abs. 2 Bachelor-SPO (7 Semester), § 23 Abs. 2 Master-SPO (3 Semester) bzw. § 19 Abs. 2 Master-SPO (4 Semester und berufsbegleitend) der HdM) einer unrichtigen oder unvollständigen ehrenwörtlichen Versicherung zur Kenntnis genommen.

(Gisela Andrea Kollotzek)
Stuttgart, den 06. Mai 2021

Gender-Erklärung

Die Verwendung des generischen Maskulinums in der vorliegenden Masterarbeit wird aus Gründen der besseren Lesbarkeit gewählt. Es werden jedoch weibliche und anderweitige Geschlechteridentitäten ausdrücklich mitgemeint.

Kurzfassung Durch die voranschreitende Digitalisierung und die steigende Anzahl von Menschen mit Behinderung hat die Relevanz der Barrierefreiheit von Informationstechnik im Laufe der letzten Jahre stark zugenommen. Mittels der Richtlinie 2016/2102 wurde die Umsetzung der Barrierefreiheit im öffentlichen Sektor für alle EU-Mitgliedsstaaten bereits verpflichtend festgelegt. Automatisierte Prüfverfahren können zwar einen ersten Eindruck über die Barrierefreiheit von Websites liefern, eine umfassende Bewertung der Barrierefreiheit ist derzeit aber nur durch manuelle Prüfmethoden möglich. Ein bekannter Typ von manuellen Methoden ist die konformitätsbasierte Prüfmethode, die den Grad der Erfüllung von definierten Anforderungen bewertet. Folglich kann abgeleitet werden, ob ein Prüfgegestand als konform oder nicht konform zu einem definierten Standard eingeordnet werden kann. Ein weiterer Typ von manuellen Methoden ist die empirisch ermittelte Prüfmethode. Sie involviert die betroffenen Benutzergruppen stärker, indem beispielsweise geschulte Prüfer die Evaluation aus Sicht der Betroffenen vornehmen.

Die im Rahmen dieser Arbeit untersuchte Forschungsfrage lautet: „Welche Unterschiede lassen sich bei konformitätsbasierten und empirisch ermittelten manuellen Prüfmethoden zur Bewertung der Barrierefreiheit von Websites feststellen?“ Ein umfassender systematischer Vergleich anhand von realen Daten, die durch Barrierefreiheitsexperten erstellt wurden, wurde bisher noch nicht durchgeführt. Um dies im Rahmen der vorliegenden Arbeit umsetzen zu können, wurde ein generischer Kriterienkatalog angefertigt, der auf dem Fachwissen diverser Barrierefreiheitsexperten basiert. Dieser Kriterienkatalog besteht aus 22 durch Experten identifizierte und gewichtete Kriterien, die durch insgesamt 41 Metriken quantifiziert werden. Der Kriterienkatalog kann in Zukunft beliebig angepasst bzw. erweitert werden, um möglichen zukünftigen Anforderungen gerecht zu werden. Mithilfe dieses erstellten Kriterienkatalogs wurden folgende zwei Prüfverfahren hinsichtlich ihrer Eignung und Wirksamkeit verglichen: Der BIK BITV-Test als eines der bekanntesten Konformitätsprüfungsverfahren in Deutschland sowie der BITV-Audit der T-Systems Multimedia Solutions GmbH als Vertreter der empirisch ermittelten Prüfmethoden.

Im Vergleich schließt der BITV-Audit anhand des definierten Kriterienkatalogs besser ab als der BIK BITV-Test. Jedoch ist zu beachten, dass es keine für alle Situationen allgemeingültige Gewichtung der Kriterien geben kann, die allen individuellen Anforderungen gerecht wird. Daher kann der Leser bei Bedarf eine individuelle Gewichtung festlegen und anhand dieser das Ergebnis des Vergleichs für beide Verfahren neu berechnen. Auf Basis der Ergebnisse wurden außerdem die Gemeinsamkeiten und Unterschiede der Prüfverfahren identifiziert. Zudem wurden die jeweiligen Schwächen und Stärken jedes Prüfverfahrens dargelegt.

Abstract Due to the advancing digitization and the increasing number of people with disabilities, accessibility of information technology has become more and more relevant over the past years. Directive 2016/2102 already enforces the implementation of accessibility in the public sector for all EU member states. Automated evaluation methods can provide a first impression of the accessibility of websites but a comprehensive assessment of accessibility is currently only possible through manual evaluation methods. A well-known type of such a manual method is the conformance-based evaluation method. It assesses the degree of compliance with the defined requirements. Following from this, it can be derived whether a website is compliant or non-compliant to the defined standard. Another type of manual method is the empirically-determined evaluation method. It involves the target user groups to a greater extent, for example, by having trained evaluators perform the evaluation from the perspective of the target group.

The research question investigated in this thesis is: „What differences can be found between conformance-based and empirically-determined manual evaluation methods for assessing website accessibility?“ Up until now, a comprehensive systematic comparison using real-world data generated by accessibility experts has not been conducted. In order to perform this task, a generic catalog of criteria was created based on the expertise of several accessibility experts. In total, the catalog of criteria consists of 22 criteria identified and weighted by experts, which are quantified using a total of 41 metrics. The catalog of criteria can be adapted or extended in the future to meet possible new requirements. With the help of this catalog of criteria, the following two evaluation methods were compared with a focus on their suitability and effectiveness: The BIK BITV-Test as one of the best known conformance-based evaluation methods in Germany, and the BITV-Audit of T-Systems Multimedia Solutions GmbH as a representative for the empirically-determined evaluation methods.

In comparison, the BITV-Audit performs better than the BIK BITV-Test based on the specified catalog of criteria. However, it should be noted that no universally valid weighting of criteria can be found for all situations which meet all individual requirements. Therefore, an individual use case based weighting can be performed and used to recalculate the result of the comparison for both methods. Based on the results, the similarities and differences between the evaluation methods were identified. In addition, the respective weaknesses and strengths of each evaluation method were described.

x

Inhaltsverzeichnis

1 Einleitung **1**

1.1 Zielsetzung 4

1.2 Aufbau der Arbeit 5

2 Methodik **7**

2.1 Literaturrecherche 7

2.2 Auswahl der Prüfverfahren 8

2.3 Identifizierung der Kriterien 9

2.4 Gewichtung der Kriterien 9

2.5 Vergleich der Prüfverfahren 9

2.6 Einschlusskriterien 10

3 Grundlagen **11**

3.1 Richtlinie 2016/2102 11

3.2 WCAG 2.1 12

3.3 EN 301 549 14

3.4 BITV 2.0 15

3.5 Zusammenhang 16

4 Verwandte Arbeiten **17**

4.1 Begrifflichkeiten 17

4.2 Vergleich verschiedener Typen von Prüfmethoden 18

4.3 Einflussfaktoren der manuellen Prüfmethoden 20

4.4 Vergleich konkreter Prüfverfahren 21

5 Bestehende Tools zur automatischen Prüfung **25**

6 Bestehende Verfahren zur manuellen Prüfung **29**

6.1 Expertenbewertung 30

6.2 Selbstbewertung 31

6.3 Auswahl der Prüfverfahren für den Vergleich 33

7 Detailbeschreibung der ausgewählten Prüfverfahren 35
7.1 BIK BITV-Test . . . 35
7.1.1 Zugrunde gelegter Standard . . . 35
7.1.2 Varianten des Prüfverfahrens . . . 36
7.1.3 Prüfvorgehen . . . 36
7.1.4 Toolunterstützung . . . 37
7.1.5 Prüferqualifikation . . . 40
7.1.6 Qualifizierung einer Organisation als Prüfstelle . . . 40
7.1.7 Prüfschritte . . . 40
7.1.8 Bewertungskategorien . . . 41
7.1.9 Qualitätssicherung . . . 41
7.1.10 Prüfbericht . . . 41
7.1.11 Veröffentlichung der Ergebnisse . . . 42
7.1.12 Prüfzeichen . . . 42
7.1.13 Preise . . . 42
7.1.14 Änderungen der Version ab März 2021 . . . 43
7.2 BITV-Audit . . . 43
7.2.1 Zugrunde gelegter Standard . . . 44
7.2.2 Varianten des Prüfverfahrens . . . 44
7.2.3 Prüfvorgehen . . . 44
7.2.4 Toolunterstützung . . . 46
7.2.5 Prüferqualifikation . . . 47
7.2.6 Prüfkriterien . . . 47
7.2.7 Bewertungskategorien . . . 48
7.2.8 Qualitätssicherung . . . 49
7.2.9 Prüfbericht . . . 49
7.2.10 Veröffentlichung der Ergebnisse . . . 50
7.2.11 Prüfzeichen . . . 50
7.2.12 Preise . . . 51

8 Erstellter Kriterienkatalog 53
8.1 Vorgehen . . . 53
8.2 Kriterien und Metriken . . . 54
8.3 Abdeckung von WCAG 2.1 AA . . . 55
8.4 Vollständigkeit . . . 55
8.5 Qualität des Prüfberichts . . . 56
8.5.1 Verfügbarkeit von Screenshots . . . 56

8.5.2 Verfügbarkeit von Markierungen in Screenshots . . . 57
8.5.3 Verfügbarkeit von dokumentierten Lösungen zu den Barrieren . . . 57
8.5.4 Anpassbarkeit der Ausführlichkeit . . . 58
8.5.5 Verständlichkeit für den Leser hinsichtlich Verstehen der Barrieren . 58
8.5.6 Barrierefreiheit des Prüfberichts . . . 59
8.6 Abdeckung von zusätzlichen Kriterien aus EN 301 549 Tab. A.1 . . . 59
8.7 Korrektheit . . . 60
8.8 Qualität der Stichprobe . . . 61
8.8.1 Umfang der Stichprobe . . . 61
8.8.2 Varianz der Stichprobe . . . 62
8.9 Qualitätssicherung . . . 62
8.9.1 Qualitätssicherung innerhalb eines Prüfauftrags . . . 63
8.9.2 Sicherstellung der Einheitlichkeit der Prüfung . . . 63
8.9.3 Sicherstellung der Qualifikation eines Prüfers . . . 64
8.10 Optionale Inputformate . . . 65
8.11 Abdeckung von zusätzlichen Kriterien aus BITV 2.0 . . . 65
8.12 Aufwand . . . 66
8.12.1 Anzahl der Prüfschritte/-kriterien . . . 66
8.12.2 Anzahl der involvierten Experten . . . 66
8.12.3 Zeitaufwand für Prüfprozess . . . 67
8.13 Abdeckung von zusätzlichen Kriterien aus WCAG 2.1 AAA . . . 67
8.14 Bandbreite an geprüften Benutzungskonstellationen . . . 68
8.14.1 Verwendete Plattformen . . . 68
8.14.2 Verwendete Browser . . . 68
8.14.3 Verwendete Hilfsmittel . . . 69
8.15 Öffentlichkeit des Prüfverfahrens . . . 69
8.16 Skalierbarkeit hinsichtlich Prüfer . . . 70
8.16.1 Einsatz von Menschen mit Behinderung als Prüfer . . . 70
8.16.2 Einsatz eines Teams von Prüfern mit unterschiedlichen Rollen . . . 71
8.16.3 Integration von Anfänger in ein Prüfteam . . . 71
8.17 Grad der Toolunterstützung . . . 72
8.17.1 Multiuserfähigkeit . . . 72
8.17.2 Plattformunabhängigkeit . . . 73
8.17.3 Automatisierung von Abläufen . . . 73
8.18 Lizenz . . . 74
8.19 Zertifikat . . . 75
8.19.1 Vorhandensein eines Prüfzeichens . . . 75

8.19.2 Veröffentlichung der Prüfergebnisse . . . 75
8.20 Potential für Automatisierung der Überprüfung . . . 76
8.21 Zukunftssicherheit für WCAG 3.0 . . . 77
8.22 Einfachheit der Prüfungsdurchführung . . . 78
8.22.1 Qualifizierungsaufwand für Prüfer . . . 78
8.22.2 Flüchtigkeitsfehler in Prüfberichten . . . 78
8.23 Organisatorische Anforderungen . . . 79
8.23.1 Finanzieller Aufwand zur Qualifikation einer Organisation . . . 79
8.23.2 Zeitlicher Aufwand zur Qualifikation einer Organisation . . . 80
8.24 Formate des Prüfberichts . . . 80

9 Vergleich der Prüfverfahren anhand des erstellten Kriterienkatalogs 81
9.1 Vorgehensbeschreibung . . . 81
9.1.1 Erstellung der Prüfberichte . . . 81
9.1.2 Aufbereitung der Prüfberichtergebnisse durch Experten . . . 83
9.1.3 Interviews . . . 86
9.2 Ergebnisse . . . 86
9.2.1 Übersicht der Top-10-Kriterien . . . 90
9.2.2 Abdeckung verschiedener Standards . . . 91
9.2.3 Qualität der Prüfung und Prüfergebnisse . . . 92
9.2.4 Effizienz der Prüfung . . . 103
9.2.5 Rahmenbedingungen des Prüfverfahrens . . . 110
9.3 Diskussion . . . 119
9.3.1 In den Vergleich eingeflossene Expertise . . . 119
9.3.2 Kriterienkatalog . . . 119
9.3.3 Gemeinsamkeiten und Unterschiede der Prüfverfahren . . . 120
9.3.4 Schwächen und Stärken der Prüfverfahren . . . 122
9.3.5 Fazit des Vergleichs . . . 124
9.3.6 Probleme nach WCAG 2.1 und aus Nutzersicht . . . 124
9.3.7 Wichtige Implikationen der Ergebnisse . . . 124
9.3.8 Limitationen . . . 125

10 Zusammenfassung 127

11 Ausblick 129

A Literaturverzeichnis 133

Abbildungsverzeichnis

2.1 Ablauf der Methodik in einzelnen Prozessschritten 8

3.1 Hierarchie der verschiedenen Ebenen von WCAG 2.1 13
3.2 Zusammenhang zwischen Richtlinie 2016/2102, BITV 2.0, EN 301 549 und WCAG 2.1 AA 16

4.1 Hierarchischer Zusammenhang zwischen den Begriffen Prüfmethode, Prüfverfahren und Prüftool 17

7.1 BIK BITV-Test: Schematische Darstellung des Prüfvorgehens 37
7.2 BIK BITV-Test: Benutzeroberfläche zum Anlegen eines neuen Prüfauftrags 38
7.3 BIK BITV-Test: Ausschnitt der Benutzeroberfläche zur Prüfung des Prüfschritts 1.1.1a - Alternativtexte für Bedienelemente 39
7.4 BIK BITV-Test: BITV-Prüfzeichen (DIAS GmbH, o.D.g) 42
7.5 BITV-Audit: Schematische Darstellung des Prüfvorgehens 45
7.6 BITV-Audit: Ausschnitt der Checkliste. Das Prüfkriterium 1.2.5 wird für drei Benutzergruppen bewertet. Das Prüfkriterium 1.2.6 wird nicht bewertet, da es dem WCAG Level AAA zugeordnet ist. 46
7.7 BITV-Audit: Menü der zusätzlichen Funktionen im Word-Dokument 47
7.8 BITV-Audit: Ausschnitt der Management-Zusammenfassung aus zwei Prüfberichten 50

9.1 Ergebnisse der zehn Kriterien mit den höchsten Gewichten 90
9.2 Ergebnisse der vier Kriterien, die die Abdeckung von Standards betreffen . 91
9.3 Ergebnisse der sechs Kriterien, die die Qualität der Prüfung und Prüfergebnisse betreffen 92
9.4 Vollständigkeit und Korrektheit der Prüfberichtergebnisse 94
9.5 Ergebnisse der drei Kriterien, die die Effizienz betreffen 104
9.6 Ergebnisse der sieben Kriterien, die die Rahmenbedingungen des Prüfverfahrens betreffen 111

Tabellenverzeichnis

5.1 Validierungswerkzeuge 26
5.2 Werkzeuge zur Prüfung von Farben und Kontrasten 26
5.3 WCAG-/BITV-Testwerkzeuge 27

6.1 Übersicht der Prüfverfahren zur Expertenbewertung 30
6.2 Übersicht der Prüfverfahren zur Selbstbewertung 32

7.1 BIK BITV-Test: Preis pro Webseite für verschiedene Komplexitätsstufen (DIAS GmbH, 2020a) 43
7.2 BITV-Audit: Problemgewichtungen und ihre Definition (T-Systems MMS, 2020b) 48
7.3 BITV-Audit: Zugänglichkeitsaussage in Abhängigkeit von Problemgewichtungen (T-Systems MMS, 2021a) 49
7.4 BITV-Audit: Preis pro Webseite für verschiedene Komplexitätsstufen (T-Systems MMS, o.D.d) 51

9.1 Überblick über die Prüfberichte der HdM und BfB 83
9.2 Überblick über aufbereitete Ergebnisse der Prüfberichte 85
9.3 Ergebnisse aller Kriterien für den BIK BITV-Test und den BITV-Audit . . 88
9.4 Ergebnisse des Kriteriums „Qualität des Prüfberichts“ 97
9.5 Ergebnisse des Kriteriums „Qualität der Stichprobe“ 99
9.6 Ergebnisse des Kriteriums „Qualitätssicherung“ 101
9.7 Ergebnisse des Kriteriums „Einfachheit der Prüfungsdurchführung“ 102
9.8 Ergebnisse des Kriteriums „Aufwand“ 105
9.9 Ergebnisse des Kriteriums „Skalierbarkeit hinsichtlich Prüfer“ 107
9.10 Ergebnisse des Kriteriums „Grad der Toolunterstützung“ 109
9.11 Ergebnisse des Kriteriums „Optionale Inputformate 112
9.12 Ergebnisse des Kriteriums „Bandbreite an geprüften Benutzungskonstellationen“ 113
9.13 Ergebnisse des Kriteriums „Öffentlichkeit des Prüfverfahrens“ 114

9.14 Ergebnisse des Kriteriums „Lizenz“ 115
9.15 Ergebnisse des Kriteriums „Zertifikat“ 116
9.16 Ergebnisse des Kriteriums „Organisatorische Anforderungen“ 118
9.17 Ergebnisse des Kriteriums „Formate des Prüfberichts“ 119
9.18 Gemeinsamkeiten und Unterschiede der Prüfverfahren 121
9.19 Schwächen und Stärken der Prüfverfahren 122

Abkürzungsverzeichnis

BfB	Bundesfachstelle für Barrierefreiheit
BGG	Behindertengleichstellungsgesetz
BIK	barrierefrei informieren und kommunizieren
BITV	Barrierefreie-Informationstechnik-Verordnung
EN 301 549	Europäische Norm 301 549
EzB	Erklärung zur Barrierefreiheit
FF	Flüchtigkeitsfehler
FN	Falsch negative Ergebnisse
FP	Falsch positive Ergebnisse
HdM	Hochschule der Medien
IKT	Informations- und Kommunikationstechnologien
KP	Korrekt positive Ergebnisse
n.b.	nicht bewertet
o.D.	ohne Datum
PAC 3	PDF Accessibility Checker Version 3
PDF	Portable Document Format
WCAG	Web Content Accessibility Guidelines
W3C	World Wide Web Consortium
T-Systems MMS	T-Systems Multimedia Solutions GmbH

1 Einleitung

> „Nicht behindert zu sein ist wahrlich kein Verdienst, sondern ein Geschenk, das jedem von uns jederzeit genommen werden kann."
> *Ehemaliger Bundespräsident Richard von Weizsäcker (1987)*

Das Leben ohne Behinderung sollte nicht als selbstverständlich angesehen werden. Viele Menschen werden früher oder später mit dieser Thematik in Berührung kommen. Dabei können die Behinderungen einen nahen Angehörigen oder auch das eigene Leben betreffen. Behinderungen können sowohl situativ, temporär oder überdauernd sein, als auch altersbedingt auftreten, wie zum Beispiel eine Verschlechterung des Sehens oder Hörens im fortgeschrittenen Lebensalter. Einer Statistik des Verband der Ersatzkassen e.V.[1] (2020a) aus dem Jahr 2019 zufolge steigt der Anteil der Leistungsempfänger der sozialen Pflegeversicherung ab einem Alter von 75 Jahren stetig an. Zudem stieg in diesem Jahr die Gesamtzahl der Leistungsempfänger der sozialen Pflegeversicherung von 3,7 auf ca. 4,0 Millionen.

In Fachkreisen gibt es eine Debatte, ob die Formulierung „Menschen mit Beeinträchtigung" der Formulierung „Menschen mit Behinderung" vorzuziehen ist. Die konsequente Verwendung des Ausdrucks „Menschen mit Behinderung" in dieser Arbeit soll ausdrücklich keine implizite Wertung hierzu abgeben. Es wurde diese Formulierung gewählt, da diese im Behindertengleichstellungsgesetz (BGG) verwendet wird. Dieses Gesetz stellt die Grundlage für die Barrierefreie-Informationstechnik-Verordnung (BITV) dar, welche eine zentrale Rolle in dieser Arbeit einnimmt. In Paragraph 3 des BGG heißt es: „Menschen mit Behinderungen im Sinne dieses Gesetzes sind Menschen, die langfristige körperliche, seelische, geistige oder Sinnesbeeinträchtigungen haben, welche sie in Wechselwirkung mit einstellungs- und umweltbedingten Barrieren an der gleichberechtigten Teilhabe an der Gesellschaft hindern können. Als langfristig gilt ein Zeitraum, der mit hoher Wahrscheinlichkeit länger als sechs Monate andauert." Im Rahmen dieser Arbeit wird diese Definition vorausgesetzt.

[1] Der Verband der Ersatzkassen e.V. ist der Interessenvertreter aller Ersatzkassen wie z.B. TK oder Barmer (vgl. Verband der Ersatzkassen e.V., 2020b).

In Deutschland gibt es keine offizielle Statistik zu Menschen mit Behinderung. Betrachtet man aber Statistiken des statistischen Bundesamt zu Personen, die als schwerbehindert eingestuft werden, wird deutlich von welchen Dimensionen auszugehen ist. Laut Sozialverbund VdK Deutschland e.V. (o.D.) gilt eine Person dann als schwerbehindert, wenn ihr Grad an Behinderung mit 50 oder höher eingestuft wird. Das Statistische Bundesamt (2020) teilte am 24. Juni 2020 mit, dass in Deutschland 7,9 Millionen schwerbehinderte Menschen leben. Insgesamt galt somit ein Anteil von 9,5% der Gesamtbevölkerung in Deutschland als schwerbehindert. Der Anteil der Personen mit Behinderung kann jedoch noch höher geschätzt werden, da einige den Grenzwert von 50 nicht überschreiten und somit nicht als schwerbehindert eingestuft werden. Das Bundesministerium für Arbeit und Soziales veröffentlichte auf der Website[2] gemeinsam-einfach-machen[3] (2019c) im Mai 2019 Zahlen und Fakten zur Lebenssituation von Menschen mit Behinderungen. Dabei wurde von einer Schätzung von 9,6 Millionen Menschen mit einer Behinderung ausgegangen, wobei zur damaligen Zeit 7,1 Millionen als schwerbehindert eingestuft wurden. Folglich hatten 11,5% der Menschen in Deutschland eine Behinderung.

Gleichzeitig spielen digitale Angebote eine immer zentralere Rolle bei der Bewältigung von Alltagsaufgaben. Beispielsweise schließen immer mehr Bankfilialen in ländlichen Regionen und verweisen auf ihr Onlineangebot. Eine Statistik (Janson, 2019), die auf Angaben der deutschen Bundesbank beruht, zeigt eine Reduktion um 24,1% der Bankfilialen von 2007 bis 2017. Durch diese Entwicklungen sind die Kunden gezwungen auf das Onlineangebot oder eine weiter entfernte Bankfiliale auszuweichen. Wenn die Onlineangebote allerdings nicht für alle zugänglich sind, werden Menschen mit Behinderung dadurch sehr stark benachteiligt. Sie sind dann gezwungen den weiten Weg auf sich zu nehmen und gegebenenfalls Hilfe hierfür in Anspruch zu nehmen.

Daraus ergibt sich eine große Notwendigkeit, dass besonders Webangebote, wie z.B. Onlinebanking-Portale, barrierefrei sind. Dies stellt nicht nur eine soziale Verantwortung dar, sondern kann vielmehr auch als eine Chance angesehen werden, wenn Menschen mit Behinderung als eine potente Kundengruppe wahrgenommen werden. So könnte es auch für bisher nicht verpflichtete Einrichtungen und Unternehmen von Vorteil sein digitale Barrierefreiheit bei der Entwicklung eigener Angebote zu berücksichtigen. In einer Forschungsstudie aus dem Jahr 2019 namens „Click-Away Pound“ (Williams und Brownlow,

[2]Eine Website stellt einen kompletten Internetauftritt dar und besteht aus mehreren Webseiten (vgl. Onlinemarketing-Praxis, o.D.b).

[3]Die Website `www.gemeinsam-einfach-machen.de` stellt Informationen rund um die UN-Behindertenrechtskonvention und den Nationalen Aktionsplan der Bundesregierung zur UN-Behindertenrechtskonvention zur Verfügung (vgl. Bundesministerium für Arbeit und Soziales, 2019a).

2020) wurde das online Einkaufserlebnis von Menschen mit Behinderung in Großbritannien erforscht und welche finanziellen Auswirkungen es für die Unternehmen hat, wenn sie diese Kundengruppe ignorieren. Die online Kaufkraft von Menschen, die auf die Barrierefreiheit von Websites angewiesen sind, wurde auf 24,8 Milliarden Pfund geschätzt. Die Studie zeigte, dass der Anteil der Menschen, die eine unzugängliche Website dann wegklicken, bei fast 70% liegt. Das entspricht umgerechnet einer Kaufkraft von 17,1 Milliarden Pfund im Jahr 2019.

Zudem gibt es zahlreiche Vorteile, die sich aus einer barrierefreien Gestaltung ergeben, von denen alle Menschen profitieren können (vgl. Schmutz, Sonderegger und Sauer, 2016). Es gibt viele Überschneidungen zwischen Barrierefreiheit und Gebrauchstauglichkeit[4] (vgl. Zimmermann, 2005). Beispielsweise ist eine Anforderung an die Barrierefreiheit, dass Linkziele eindeutig formuliert werden sollen (vgl. W3C, 2018). Analog dazu existiert eine Anforderung an die Gebrauchstauglichkeit, dass Linkbezeichnungen kurz und eindeutig formuliert werden sollen (vgl. ISO, 1999). Diese Schnittmenge führt dazu, dass eine hohe Barrierefreiheit auch zu einer verbesserten Gebrauchstauglichkeit für alle Benutzer führt.

Aufgrund der europäischen Richtlinie 2016/2102 (Europäisches Parlament und Rat der europäischen Union, 2016) muss es seit September 2018 gesetzliche Regelungen in allen EU-Mitgliedsstaaten geben, die staatliche Einrichtungen verpflichten die Barrierefreiheit für Websites und mobile Anwendungen zu gewährleisten. Zur Bewertung der Barrierefreiheit können automatische Prüfmethoden genutzt werden, es können aber nicht alle relevanten Anforderungen damit hinreichend geprüft werden (vgl. Vigo, Brown und Conway, 2013). Beispielsweise ist gesetzlich festgelegt, dass Textalternativen für Bilder bereitgestellt werden müssen, um so die Information auch für Benutzer von Screenreadern zugänglich zu machen. Automatische Prüfmethoden können zwar erkennen, ob eine Alternative in Textform bereitgestellt wird, aber nicht, ob der Text wirklich eine sinnvolle Beschreibung darstellt (vgl. BIK für alle, o.D.). Die Sinnhaftigkeit kann nur von Menschen bewertet werden. Eine Studie von Burkard, Zimmermann und Schwarzer (2021) zeigt, dass mit diversen getesteten automatischen Prüftools nur 15 bis 40% aller bestehenden Barrierefreiheitsprobleme in den untersuchten Prüfgegenständen gefunden werden können. Dadurch wird deutlich, dass keins der getesteten Prüftools die Barrierefreiheit automatisch vollständig prüfen kann. Im Gegenteil, es werden im Schnitt sogar ca. zwei Drittel aller notwendigen Prüfschritte gar nicht geprüft. Das heißt, die automatischen Prüfmethoden können nur einen ersten Eindruck über den Stand der Barrierefreiheit bieten. Sie können allerdings nicht feststellen, ob

[4] engl. usability

die Website tatsächlich barrierefrei ist. Derzeit können nur manuelle Prüfmethoden eine umfassende Prüfung auf Barrierefreiheit gewährleisten. Diese Prüfverfahren sind allerdings auch mit einem großen zeitlichen Aufwand und der notwendigen Qualifizierung der Prüfer verbunden. So sind die Erstellung von barrierefreien Inhalten sowie die Zertifizierung der Barrierefreiheit ebenfalls mit einem großen Aufwand und damit einhergehenden hohen Kosten verbunden. Derzeit gibt es noch wenig Forschung im Bereich der manuellen im Vergleich zu den automatischen Prüfmethoden. Jedoch ist die Weiterentwicklung der manuellen Prüfmethoden essentiell, um mehr barrierefreie Webangebote bereitstellen zu können.

1.1 Zielsetzung

Der Fokus dieser Arbeit liegt auf der Prüfung der Barrierefreiheit von Webangeboten. Mit Webangeboten sind alle Formen von Websites und Webanwendungen[5] gemeint. Außerdem zählen auch Dateiformate wie z.B. PDF-Dokumente, die über das Internet geteilt werden können, dazu. Nicht eingeschlossen sind digitale Angebote wie mobile Anwendungen (auch App genannt) oder Software. Die zu untersuchende Forschungsfrage lautet:

> „**Welche Unterschiede lassen sich bei konformitätsbasierten und empirisch ermittelten manuellen Prüfmethoden zur Bewertung der Barrierefreiheit von Websites feststellen?**“

Das Ziel der Arbeit ist die Analyse dieser beiden Arten von manuellen Prüfmethoden hinsichtlich ihrer Eignung und Wirksamkeit.

Es handelt sich bei beiden Arten um manuelle Prüfmethoden. Sie unterscheiden sich allerdings darin, dass die konformitätsbasierten Prüfmethoden den Grad der Erfüllung von definierten Anforderungen bewerten. Folglich kann abgeleitet werden, ob der Prüfgegenstand in seiner Gesamtheit als konform oder nicht konform zu einem definierten Standard eingeordnet werden kann. Bei den empirisch ermittelten Prüfmethoden werden die betroffenen Benutzergruppen (z.B. sehbeeinträchtigte Benutzer) stärker involviert, indem beispielsweise entsprechend geschulte Prüfer die Prüfung auf Barrierefreiheit aus der Perspektive dieser Benutzergruppen vornehmen. Zudem kann beurteilt werden wie gravierend die gefundenen Probleme für die jeweiligen Benutzergruppen sind.

[5] Webanwendungen haben handlungsorientierte, interaktive Elemente im Gegensatz zu Websites, die meist einen informierenden Charakter haben (vgl. 1&1 IONOS SE, 2021).

Zur Durchführung eines Vergleichs werden zwei Vertreter der beiden Arten von manuellen Prüfmethoden ausgewählt: Der BIK BITV-Test (konformitätsbasiert) und der BITV-Audit von T-Systems Multimedia Solutions GmbH (empirisch ermittelt). Ferner werden Kriterien von Experten definiert und gewichtet. Anhand dieses Kriterienkatalogs wird ein Vergleich durchgeführt. Es werden die Gemeinsamkeiten und Unterschiede sowie die Schwächen und Stärken der Prüfverfahren identifiziert. So soll diese Arbeit eine Hilfestellung bei der Entscheidung bieten, welches der beiden Prüfverfahren besser geeignet ist. Da es keine allgemeingültige Gewichtung geben kann, welche den Anforderungen aller individuellen Situationen gerecht wird, kann der unter der gegebenen Expertengewichtung errechnete Gesamtscore der Prüfverfahren nur als erster Anhaltspunkt dienen. Der Leser kann allerdings auch eine eigene Gewichtung festlegen, die der jeweiligen Situation entspricht und so eine fundierte Entscheidung treffen.

1.2 Aufbau der Arbeit

Die Arbeit gliedert sich wie folgt:

Kapitel 2 beschreibt die zur Untersuchung der Forschungsfrage angewendete Methodik.

Kapitel 3 zeigt die notwendigen Grundlagen auf. Hierbei werden die rechtlichen Rahmenbedingungen betrachtet und in Zusammenhang gebracht.

Kapitel 4 beleuchtet den derzeitigen Forschungsstand im Bereich der manuellen Prüfmethoden.

Kapitel 5 geht auf die bereits bestehenden Tools zur automatischen Prüfung der Barrierefreiheit von Websites ein.

Kapitel 6 stellt eine Übersicht der bereits bestehenden Verfahren zur manuellen Prüfung der Barrierefreiheit von Websites vor, um dem Leser so einen umfassenden Überblick über beide Arten von Prüfmethoden zu ermöglichen. Im Nachfolgenden liegt der Fokus ausschließlich auf den Verfahren zur manuellen Prüfung.

Kapitel 7 nimmt eine detaillierte Betrachtung der beiden zum Vergleich ausgewählten manuellen Prüfverfahren vor, um beim Leser ein Grundverständnis für diese zu schaffen.

Kapitel 8 präsentiert die durch die Experten festgelegten Kriterien sowie deren Gewichtung. Außerdem werden hier auch die jeweiligen Metriken für die einzelnen Kriterien näher erläutert.

Kapitel 9 geht detailliert auf das Vorgehen des anhand der Kriterien durchgeführten Vergleichs ein. Anschließend werden die Ergebnisse zusammengefasst und diskutiert.

Kapitel 10 fasst die Ergebnisse dieser Arbeit zusammen und geht insbesondere auf die neuen Erkenntnisse ein.

Kapitel 11 gibt einen Ausblick über mögliche künftige Forschungsthemen in diesem Bereich.

2 Methodik

Zur Untersuchung der zentralen Forschungsfrage dieser Arbeit wird ein Mixed-Methods-Ansatz genutzt. Das heißt, es werden sowohl quantitative als auch qualitative Forschungsmethoden verwendet, sodass die Forschungsfrage aus verschiedenen Blickwinkeln betrachtet werden kann. Der quantitative Vergleich der Prüfverfahren wird durch qualitative Experteninterviews ergänzt. Dadurch soll eine umfassende Untersuchung der Forschungsfrage gewährleistet werden.

Abbildung 2.1 zeigt das strukturierte methodische Vorgehen dieser Arbeit. Um die Grundlage für den anschließenden Vergleich zu schaffen, wurde eine ausführliche Literaturrecherche durchgeführt. Auf Basis der Recherchearbeit wurden die Prüfverfahren ausgewählt, die im Vergleich gegenübergestellt werden. Die für den Vergleich notwendigen Kriterien wurden durch eine Fokusgruppe bestehend aus Barrierefreiheitsexperten identifiziert und gewichtet. So kann ein fundierter Vergleich der Prüfverfahren erfolgen.

In den nachfolgenden Abschnitten wird aufgezeigt, wie diese Aspekte im Rahmen dieser Arbeit umgesetzt wurden. Außerdem werden in Abschnitt 2.6 die Einschlusskriterien dargelegt.

2.1 Literaturrecherche

Die Literaturrecherche beruht auf Rechtsvorschriften, öffentlich zugänglichen und vertrauenswürdigen Internetquellen sowie veröffentlichten wissenschaftlichen Artikeln. Zuerst wurde die der Barrierefreiheit im digitalen Raum zugrunde liegende Gesetzgebung näher betrachtet und in Zusammenhang gebracht. Anschließend wurde der momentane Forschungsstand im Bereich der manuellen Prüfmethoden zur Bewertung der Barrierefreiheit von Websites beleuchtet. Die Literaturrecherche wurde abgeschlossen mit einer umfassenden Recherche zu bereits bestehenden Methoden zur manuellen und automatischen Prüfung. Der Überblick über automatische Prüfmethoden wird zur Vollständigkeit

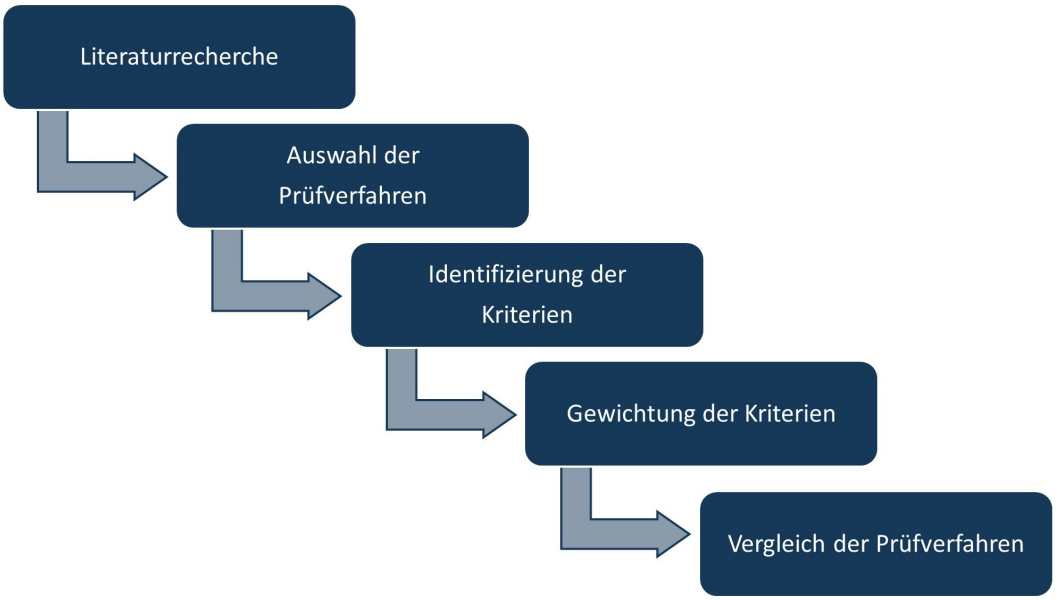


Abbildung 2.1: Ablauf der Methodik in einzelnen Prozessschritten

gegeben, der Fokus dieser Arbeit liegt jedoch auf den manuellen Prüfmethoden. Die Ergebnisse der Literaturrecherche sind in Kapitel 3, Kapitel 4, Kapitel 5 sowie Kapitel 6 zu finden.

2.2 Auswahl der Prüfverfahren

Auf Basis der in der Literaturrecherche erstellten Übersicht über manuelle Prüfmethoden wurden zwei Prüfverfahren ausgewählt, um in einer Analyse gegenübergestellt zu werden. Als Vertreter der konformitätsbasierten manuellen Prüfmethoden wurde der BIK BITV-Test ausgewählt, da es sich um ein in Deutschland etabliertes Prüfverfahren handelt. Wenn der BIK BITV-Test durch eine vom BIK-Prüfverbund anerkannte Prüfstelle angewendet wird, kann dadurch eine Zertifizierung der Barrierefreiheit durchgeführt werden. Um dem ein vergleichbares Prüfverfahren der empirisch ermittelten manuellen Prüfmethoden gegenüberzustellen, wurde der BITV-Audit der T-Systems Multimedia Solutions GmbH (T-Systems MMS) ausgewählt. Es handelt sich um ein akkreditiertes Prüfverfahren, das ebenfalls eine Zertifizierung der Barrierefreiheit vornehmen kann. Eine detaillierte Beschreibung der Prüfverfahren ist in Kapitel 7 aufgeführt.

2.3 Identifizierung der Kriterien

Um einen fundierten Vergleich der beiden Prüfverfahren durchführen zu können, muss eine Auswahl an Kriterien getroffen werden. Basierend auf der Literaturrecherche und einem Brainstorming mit den Betreuern dieser Arbeit wurden einige Kriterien definiert. Diese Vorauswahl wurde in einer Fokusgruppe bestehend aus Barrierefreiheitsexperten diskutiert, überarbeitet und mit weiteren Kriterien ergänzt. Eine detaillierte Vorgehensbeschreibung und die vollständige Übersicht der gefundenen Kriterien befinden sich in Kapitel 8.

2.4 Gewichtung der Kriterien

Im Anschluss an die Expertenrunde vergab jeder Barrierefreiheitsexperte seine persönliche Gewichtung. Die Gewichtungen wurden jeweils normalisiert und anschließend gemittelt. Eine detaillierte Vorgehensbeschreibung und die Übersicht der festgelegten Gewichte für die Kriterien sind in Kapitel 8 zu finden.

2.5 Vergleich der Prüfverfahren

Der Vergleich der Prüfverfahren beruht sowohl auf den derzeit aktuellen Verfahrensbeschreibungen als auch auf eigens für den Vergleich erstellten Prüfberichten. Ein qualifizierter Prüfer, der beide Prüfverfahren beherrscht, prüft Websites mit beiden Prüfverfahren. Dadurch entsteht eine Datenbasis, die für jede Website zwei Prüfberichte des gleichen Prüfers bereitstellt. Anschließend wurden die Prüfberichte in einem Workshop mit Barrierefreiheitsexperten aufbereitet. Für Kriterien, deren Bewertungen sich nicht anhand der Verfahrensbeschreibung oder Prüfberichte ergeben, wurden zusätzlich Experteninterviews durchgeführt.

Die Methode des Interviews wurde einer Umfrage vorgezogen, da so eine hohe Rücklaufquote erwartet werden konnte. Durch die direkte Kommunikation hatten die Experten zudem die Möglichkeit, gezielte Nachfragen stellen zu können und über die Bewertung hinausgehende Gedanken zu äußern. Ein Nachteil dieser Methode ist der hohe Zeitaufwand im Vergleich zu einer Umfrage. Aufgrund der Tatsache, dass es sich hierbei um die Erfassung der Meinung weniger Experten handelt, hielt sich der Zeitaufwand im Rahmen.

Ebenso war die Gefahr, dass erwünschte Antworten bei einer direkten Befragung gegeben werden könnten, eher gering, da die Interviewerin keine direkte Verbindung zu einer Organisation der Prüfverfahren hatte.

Auf Basis der gesammelten Daten sowie des erstellten Kriterienkatalogs wurde der Vergleich der Prüfverfahren durchgeführt. So kann auf Grundlage der gefundenen Ergebnisse eine Aussage zu den Gemeinsamkeiten und Unterschieden der beiden Prüfverfahren gegeben werden. Darüber hinaus werden die jeweiligen Schwächen und Stärken der einzelnen Prüfverfahren aus den Ergebnissen abgeleitet. Außerdem ermöglicht der detailliert dokumentierte Vergleich der Prüfverfahren dem Leser das Anwenden einer individuellen Gewichtung. Eine detaillierte Vorgehensbeschreibung sowie die gefundenen Ergebnisse werden in Kapitel 9 dargestellt und diskutiert.

2.6 Einschlusskriterien

Folgende Einschlusskriterien waren ausschlaggebend, damit die Daten in die Forschung einbezogen wurden: Es wurden nur Gewichtungen berücksichtigt, die vollständig und korrekt ausgefüllt wurden. Zur Evaluation der Prüfverfahren wurden ausschließlich die eigens für den Vergleich erstellten Prüfberichte sowie die zu derzeit aktuellen Verfahrensbeschreibungen genutzt. Außerdem wurden lediglich Interviews mit für das jeweilige Prüfverfahren qualifizierten Prüfern durchgeführt. Zudem wurden ausnahmslos Interviews berücksichtigt, bei denen alle Fragestellungen vollständig beantwortet wurden.

3 Grundlagen

Zuerst wird in Abschnitt 3.1 die Richtlinie 2016/2102 betrachtet, da diese alle Mitgliedsstaaten der Europäischen Union dazu verpflichtet hat, den barrierefreien Zugang zu Websites und mobilen Anwendungen bei öffentlichen Stellen in nationales Recht umzusetzen. Außerdem werden in den nachfolgenden Abschnitten die in der Richtlinie 2016/2102 genannten Richtlinien für Webinhalte, die Europäische Norm 301 549 sowie die BITV 2.0 näher beleuchtet und in Abschnitt 3.5 deren Zusammenhang herausgearbeitet. Somit erhält der Leser einen umfassenden Einblick in die derzeit geltenden rechtlichen Rahmenbedingungen.

3.1 Richtlinie 2016/2102

Die Richtlinie 2016/2102 (Europäisches Parlament und Rat der europäischen Union, 2016) über den barrierefreien Zugang zu den Websites und mobilen Anwendungen öffentlicher Stellen wurde vom europäischen Parlament und Rat am 26. Oktober 2016 erlassen und im Dezember 2016 im Amtsblatt der Europäischen Union veröffentlicht. Diese Richtlinie wird auch Web Accessibility Directive genannt. Hierbei handelt es sich um eine EU-Richtlinie, die nicht unmittelbar wirksam wurde, sondern innerhalb von zwei Jahren von den Mitgliedsstaaten der Europäischen Union in nationales Recht umgesetzt werden musste. Die Richtlinie fordert eine Mindestharmonisierung, d.h. die Mitgliedsstaaten müssen mindestens alle in der Richtlinie aufgeführten Anforderungen erfüllen, können aber darüber hinaus noch weitere hinzufügen. Ein Beispiel für eine solche Anforderung ist: „Die Mitgliedstaaten stellen sicher, dass öffentliche Stellen eine detaillierte, umfassende und klare Erklärung zur Barrierefreiheit über die Vereinbarkeit ihrer Websites und mobilen Anwendungen mit dieser Richtlinie bereitstellen und diese regelmäßig aktualisieren.“ (Europäisches Parlament und Rat der europäischen Union, 2016). In Deutschland wurde die Richtlinie im Juli 2018 im BGG umgesetzt. Zudem wurde im Mai 2019 die BITV 2.0 überarbeitet.

Die Richtlinie 2016/2102 verpflichtet öffentliche Stellen aller Ebenen (Bundes-, Landes- und Kommunalebene) zur Bereitstellung von barrierefreien Webangeboten. Unter Webangeboten werden Websites, Intranet, mobile Anwendungen sowie Dateiformate von Büroanwendungen (z.B. PDF-Dokumente) verstanden. Von dieser Pflicht sind nur wenige Ausnahmen befreit, wie zum Beispiel Webangebote von öffentlich-rechtlichen Rundfunkanstalten. Die Richtlinie nimmt an, dass Inhalte von Webangeboten, welche harmonisierten Normen entsprechen, die Barrierefreiheitsanforderungen erfüllen. Für Websites ist die Europäische Norm 301 549 (EN 301 549) festgelegt. Darüber hinaus werden noch weitere Anforderungen definiert, wie zum Beispiel, dass öffentliche Stellen eine Erklärung zur Barrierefreiheit (EzB) ihrer Webangebote zur Verfügung stellen müssen.

3.2 Web Content Accessibility Guidelines (WCAG)

Die Richtlinien für barrierefreie Webinhalte (WCAG) stellen weltweit anerkannte Empfehlungen dar, um Inhalte von Webseiten[1] barrierefrei zu gestalten. Durch das Befolgen der Richtlinien gewährleisten Website-Betreiber, dass ihre Internetangebote von einer größeren Gruppe von Menschen mit Behinderungen barrierefrei genutzt werden können. Die derzeit aktuellste Version 2.1 wurde am 5. Juni 2018 vom World Wide Web Consortium[2] (W3C) veröffentlicht (vgl. W3C, 2018). Bei diesen Richtlinien handelt es sich um Empfehlungen, diese sind als solche nicht verpflichtend.

Wie in Abbildung 3.1 dargestellt, sind die WCAG hierarchisch aufgebaut. Den Prinzipien unterstehen die Richtlinien. Diesen wiederum werden die Erfolgskriterien zugeordnet. In Abbildung 3.1 ist beispielhaft das erste Prinzip „wahrnehmbar“ mit seinen ersten beiden Richtlinien und jeweils einem Erfolgskriterium abgebildet.

WCAG umfasst die folgenden vier Prinzipien: wahrnehmbar, bedienbar, verständlich und robust. Diese stellen die Grundlage der Barrierefreiheit dar. Ist eines der Prinzipien verletzt, gilt der Webinhalt nicht als barrierefrei. Unter „wahrnehmbar“ ist zu verstehen, dass die Webinhalte vom Benutzer mit einem seiner Sinne erfasst werden können und kein Inhalt verborgen bleibt. Das Prinzip „bedienbar“ besagt, dass es keine Interaktion gibt, die vom Benutzer nicht ausgeführt werden kann. Damit ist beispielsweise eine ausschließlich mit der Maus auszuführende Interaktion gemeint, die der Benutzer aufgrund seiner

[1] Eine Webseite ist eine Seite einer Website (vgl. Onlinemarketing-Praxis, o.D.a).
[2] Das World Wide Web Consortium ist eine internationale Gemeinschaft, die an der Entwicklung von Webstandards arbeitet (vgl. W3C, o.D.a).

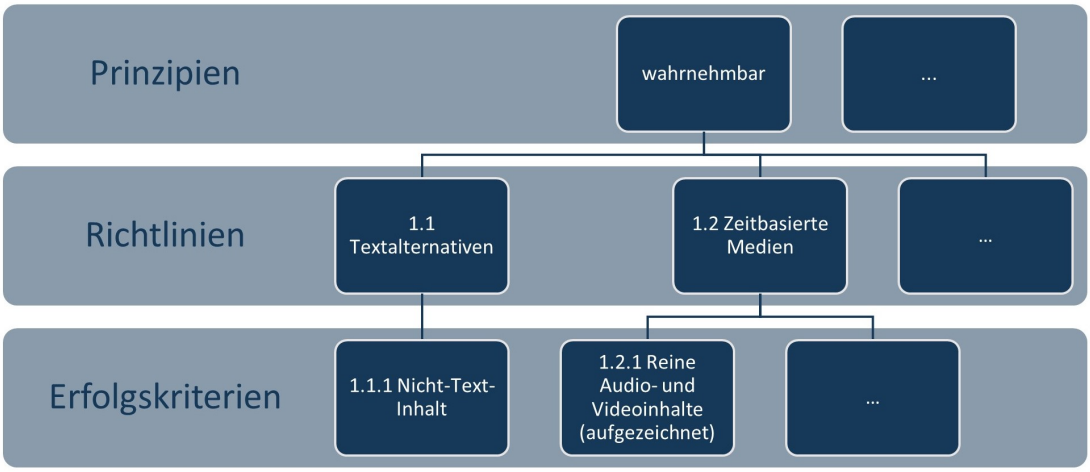


Abbildung 3.1: Hierarchie der verschiedenen Ebenen von WCAG 2.1

Behinderung nicht verwenden kann. Das Prinzip „verständlich“ beinhaltet, dass die Informationen und Interaktionsmöglichkeiten vom Benutzer verstanden werden können. Somit darf keine besondere Kenntnis vorausgesetzt werden, die für das Verstehen der Interaktion sowie der Inhalte notwendig ist. Das letzte Prinzip „robust“ bezieht sich darauf, dass die Webinhalte von einer Vielzahl von Nutzeragenten wie Webbrowsern oder assistiven Technologien, zum Beispiel Screenreadern, verarbeitet werden können, auch wenn diese sich weiterentwickeln (vgl. W3C, o.D.b).

Eine Ebene unter den Prinzipien befinden sich die 13 Richtlinien. Diese sind als wesentliche Ziele zu verstehen und jeweils einem Prinzip zugeordnet. Sie sind wichtig, um die ihnen zugeordneten Erfolgskriterien besser zu verstehen und testen zu können. Die Richtlinien sowie Prinzipien selbst sind abstrakter als die Erfolgskriterien und daher nicht testbar (vgl. W3C, 2018).

In der Version 2.1 (W3C, 2018) gibt es insgesamt 78 Erfolgskriterien, die im Gegensatz zu den Richtlinien testbar sind. Jedes Erfolgskriterium ist einer der insgesamt drei Konformitätsstufen zugeordnet. 30 Erfolgskriterien gehören der niedrigsten Stufe (A) und 20 der mittleren Stufe (AA) an. Die höchste Konformitätsstufe (AAA) beinhaltet die übrigen 28 Erfolgskriterien. Diese Konformitätsstufen entsprechen unterschiedlichen Niveaus bei der Umsetzung von Barrierefreiheit. Soll eine Website der Konformitätsstufe AA entsprechen, müssen hierfür alle Erfolgskriterien der Stufe A und der Stufe AA erfüllt sein.

Eine Aktualisierung der WCAG 2.1 auf Version 2.2 ist für das Jahr 2021 geplant (vgl. W3C, 2020). Momentan wird bereits an einer neuen Version namens „W3C Accessibi-

lity Guidelines 3.0“ gearbeitet. Die Veröffentlichung eines ersten Arbeitsstands ist für Ende 2021 geplant. Die finale Version ist für Ende 2022 angekündigt (vgl. W3C, 2021). Die kontinuierliche Weiterentwicklung des Standards ist notwendig, um an die immer neu aufkommenden technischen Möglichkeiten und daraus resultierenden Benutzungsszenarien anzuschließen. Im Rahmen dieser Arbeit wird ausschließlich Bezug zu WCAG 2.1 (W3C, 2018) genommen, da diese Version die derzeit gültige ist.

3.3 EN 301 549

Die EN 301 549 V2.1.2 (Europäisches Institut für Telekommunikationsnormen, 2018) legt Anforderungen zur Barrierefreiheit von Informations- und Kommunikationstechnologien (IKT) fest. Diese harmonisiert mit der Richtlinie 2016/2102 (Europäisches Parlament und Rat der europäischen Union, 2016). In der Richtlinie 2016/2102 wurde festgeschrieben, dass die EN 301 549 angewendet werden muss und somit gilt die Konformität mit den darin genannten Anforderungen. In der EN 301 549 werden alle Bereiche der IKT abgedeckt: IKT mit Zwei-Wege-Sprachkommunikation, IKT mit Videofunktionen, Hardware, Web, Nicht-Web-Dokumente, Software, Dokumentation und Supportleistungen sowie IKT für den Zugang zu Relais- oder Notdiensten.

Das neunte Kapitel „Web“ der EN 301 549 (Europäisches Institut für Telekommunikationsnormen, 2018) definiert alle Anforderungen für Websites. Dabei orientiert sich die EN 301 549 hinsichtlich der Webinhalte stark an WCAG 2.1 AA. Somit müssen alle Erfolgskriterien der Stufe A und der Stufe AA erfüllt sein. Im Anhang A der EN 301 549 wird der Zusammenhang zwischen der Norm und den Anforderungen der Richtlinie 2016/2102 genau definiert. In Tabelle A.1 werden auch weitere Anforderungen für Webinhalte genannt, die über die Anforderungen der WCAG 2.1 AA hinausgehen. Folglich sind alle Anforderungen der WCAG 2.1 AA in der EN 301 549 enthalten, wobei zusätzlich noch weitere Anforderungen für Webinhalte definiert werden. Zum Beispiel besagt eine Anforderung: Wenn die IKT biologische Merkmale verwendet, darf sie sich nicht auf die Verwendung eines bestimmten biologischen Merkmals als einziges Mittel zur Benutzeridentifikation oder zur Kontrolle der IKT stützen (vgl. Europäisches Institut für Telekommunikationsnormen, 2018, S.28).

Die momentan geltende Version V2.1.2 wurde im August 2018 veröffentlicht. Die Herausgeber sind das Europäische Institut für Telekommunikationsnormen, das Europäische Komitee für Normung und das Europäische Komitee für elektrotechnische Normung (vgl. Europäisches Institut für Telekommunikationsnormen, 2018). Derzeit gibt es bereits einen

Arbeitsstand der neuen Version V3.2.1 (vgl. Europäisches Institut für Telekommunikationsnormen, 2020). Im Rahmen dieser Arbeit wird immer Bezug auf den Arbeitsstand der neuen Version V3.2.1 genommen, da dieser zeitnah auch offizielle Gültigkeit haben soll. An dieser Stelle soll jedoch deutlich gemacht werden, dass es sich bei der Version 3.2.1 zum jetzigen Zeitpunkt nicht um die offizielle Europäische Norm handelt, sondern um den Arbeitsstand der Europäischen Norm.

3.4 BITV 2.0

Die BITV 2.0 (Bundesministerium für Arbeit und Soziales, 2019b) stellt die Verordnung zur Schaffung barrierefreier Informationstechnik nach dem BGG dar. Die Version 2.0 löste im Jahr 2011 die davor geltende BITV ab und wurde zuletzt im Mai 2019 geändert. Durch diese letzten Änderungen setzt sie die Vorgaben der Richtlinie 2016/2102 um, die 2018 nicht im BGG aufgenommen wurden (vgl. Bundesministerium für Arbeit und Soziales, 2019b).

Umgesetzt wurde dies im Paragraph 3 Absatz 1 der Verordnung (Bundesministerium für Arbeit und Soziales, 2019b). Dort heißt es, dass die Angebote und Dienste barrierefrei gestaltet werden müssen. In diesem Kontext wird auch definiert, dass die vier Prinzipien der WCAG 2.1 erfüllt sein müssen. Im Paragraph 3 Absatz 2 steht, dass die Angebote und Dienste harmonisierten Normen entsprechen müssen und diese harmonisierten Normen im Amtsblatt der Europäischen Union genannt worden sind. Diese Europäische Norm ist derzeit die EN 301 549.

Um der Richtlinie 2016/2102 zu entsprechen, gibt es darüber hinaus zusätzliche Anforderungen in der BITV 2.0 (Bundesministerium für Arbeit und Soziales, 2019b), die über die EN 301 549 hinausgehen. Dazu zählt die EzB. Sie gibt den Benutzern Informationen über den momentanen Stand der Barrierefreiheit der Website. Hierzu muss eine tatsächliche Prüfung der Website durchgeführt werden. Diese kann durch die öffentliche Stelle selbst oder einen Dritten vorgenommen werden. Die EzB muss von jeder Webseite einer Website erreichbar sein und mindestens jährlich aktualisiert werden. Bei wesentlichen Änderungen der Website sollte diese unmittelbar aktualisiert werden. Zudem muss es einen Feedback-Mechanismus für die Benutzer geben, über den nicht-barrierefreie Inhalte direkt gemeldet werden können. Diese Funktion muss ebenfalls von allen Webseiten einer Website erreichbar sein. Informationen zu wesentlichen Inhalten der Website sowie Hinweise zur Navigation müssen in Deutscher Gebärdensprache sowie in Leichter Sprache bereitgestellt werden. Außerdem müssen wesentliche Inhalte der EzB sowie Hinweise auf weitere

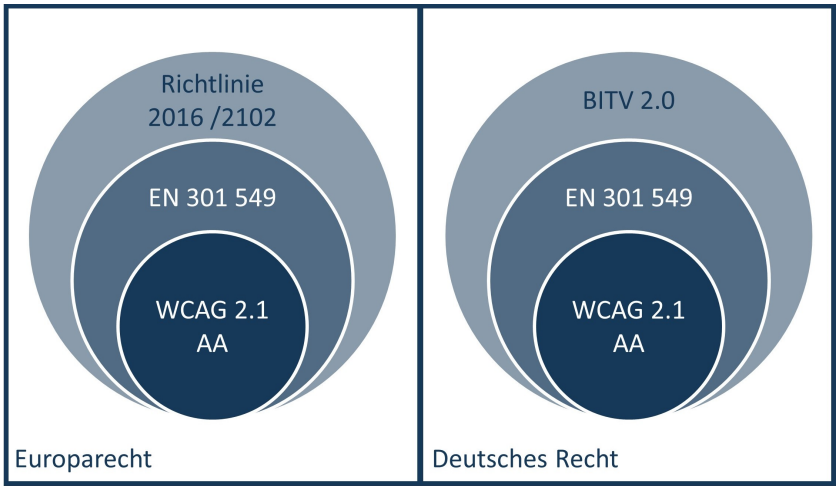


Abbildung 3.2: Zusammenhang zwischen Richtlinie 2016/2102, BITV 2.0, EN 301 549 und WCAG 2.1 AA

zur Verfügung stehende Inhalte in Deutscher Gebärdensprache und in Leichter Sprache verfügbar sein.

3.5 Zusammenhang

Abbildung 3.2 visualisiert den Zusammenhang der bisher beschriebenen Richtlinien, der Norm und der Verordnung. Die Richtlinie 2016/2102 verweist auf die EN 301 549, beinhaltet jedoch zusätzliche Anforderungen. Die EN 301 549 umfasst alle Anforderungen der WCAG 2.1 AA, wobei darüber hinaus weitere Anforderungen definiert werden. Eine Website ist dann als barrierefrei einzuschätzen, wenn alle in der Richtlinie 2016/2102 festgelegten Anforderungen erfüllt sind.

Da es sich bei der Richtlinie 2016/2102 um eine europäische Richtlinie handelt, musste diese im nationalen Recht der Mitgliedsstaaten umgesetzt werden. In Deutschland wurde dies zusätzlich zur Änderung des BGG durch die Aktualisierung der BITV 2.0 erreicht. Wie die Richtlinie 2016/2102 verweist auch die BITV 2.0 auf die EN 301 549. Außerdem wurden die zusätzlichen Anforderungen der Richtlinie 2016/2102 aufgenommen. Durch den Verweis auf EN 301 549 sind alle WCAG 2.1 AA Anforderungen sowie die zusätzlichen Anforderungen der EN 301 549 abgedeckt. Eine Website ist dann als barrierefrei einzuschätzen, wenn sie alle in der BITV 2.0 genannten Anforderungen erfüllt.

4 Verwandte Arbeiten

Im Folgenden werden wichtige Forschungsarbeiten zusammengefasst, sodass der Leser einen Eindruck des aktuellen Forschungsstands im Bereich der manuellen Prüfmethoden erhält. In Abschnitt 4.1 werden die Begrifflichkeiten erläutert, damit die Unterschiede zwischen Prüfmethoden, Prüftools und Prüfverfahren deutlich werden. In Abschnitt 4.2 werden Arbeiten vorgestellt, die sich auf den Vergleich der verschiedenen Typen von Prüfmethoden konzentrieren. Anschließend werden in Abschnitt 4.3 Arbeiten zusammengefasst, die sich mit wichtigen Einflussfaktoren auf die manuellen Prüfmethoden beschäftigen. Ferner werden in Abschnitt 4.4 Forschungsarbeiten beleuchtet, die sich auf den Vergleich konkreter manueller Prüfverfahren fokussieren.

4.1 Begrifflichkeiten

Abbildung 4.1 zeigt die Zusammenhänge der verwendeten Begrifflichkeiten. Im Rahmen dieser Arbeit werden die Begriffe Prüfmethode, Prüfverfahren und Prüftool wie folgt verwendet: Die Prüfmethode umfasst sowohl manuelle als auch automatische Methoden zur

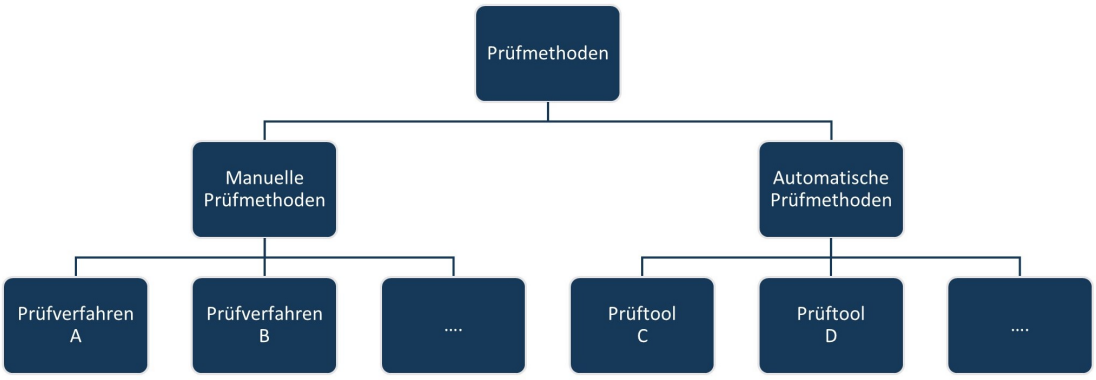


Abbildung 4.1: Hierarchischer Zusammenhang zwischen den Begriffen Prüfmethode, Prüfverfahren und Prüftool

Bewertung der Barrierefreiheit von Websites. Dieser Begriff wird verwendet, wenn auf einer abstrakten Ebene über die verschiedenen Typen von Prüfmethoden diskutiert wird. Ein Prüfverfahren ist eine konkrete Instanz einer manuellen Prüfmethode. Dem gegenüber steht das Prüftool, welches eine konkrete Instanz einer automatischen Prüfmethode darstellt.

4.2 Vergleich verschiedener Typen von Prüfmethoden

Lang (2004) betrachtet drei verschiedene Typen von Prüfmethoden zur Bewertung der Barrierefreiheit von Websites hinsichtlich ihrer Effektivität mittels einer Literaturrecherche: automatische sowie manuelle Prüfungen und Benutzertests mit Menschen mit Behinderung. Es zeigt sich, dass keine der untersuchten Prüfmethoden alle Barrierefreiheitsprobleme identifizieren kann. Lang kommt zu dem Schluss, dass eine Kombination der drei Methodentypen der effektivste Weg ist, um die größtmögliche Anzahl an Barrierefreiheitsproblemen zu finden und so die Barrierefreiheit verbessern zu können. Für Organisationen, deren finanzielle und zeitliche Mittel begrenzt sind, wird jedoch eine Kombination aus einer automatischen und manuellen Prüfung empfohlen, um eine große Anzahl an diversen Barrierefreiheitsproblemen erkennen zu können. Es wird auf einen Benutzertest mit Menschen mit Behinderung verzichtet, um Kosten und Zeit zu sparen. Dieser Vergleich basiert ausschließlich auf Literaturrecherche und vergleicht daher die Typen der Prüfmethoden auf konzeptioneller Ebene. Somit werden Aspekte, die sich bei einer praktischen Durchführung ergeben, außer Acht gelassen.

In der Studie von Mankoff, Fait und Tran (2005) werden ebenfalls die bereits von Lang genannten Prüfmethoden miteinander verglichen. Um diesen Vergleich durchführen zu können, wurde jedoch für jede Prüfmethode ein konkretes Prüfverfahren bzw. Prüftool ausgewählt: eine Konformitätsprüfung[1] zu den WCAG 1.0 mit Screenreader und ohne Screenreader, ein Prüftool (namens Bobby 4.0) sowie eine remote Prüfung durch blinde Benutzer. Dafür verwendeten die blinden Benutzer ihren eigenen PC und teilten die gefundenen Barrieren per Mail mit. Der Vergleich beschränkt sich auf die folgenden Kriterien: die Wirksamkeit, Validität und Anzahl der gefundenen Probleme. Es zeigt sich, dass die durch mehrere Prüfer durchgeführte Konformitätsprüfung mit Screenreader am erfolgreichsten ist, verschiedene Klassen von Problemen zu finden. Mankoff et al. geben

[1] Konformitätsprüfung ist gleichbedeutend mit der konformitätsbasierten Prüfmethode wie sie in Abschnitt 1.1 beschrieben wird.

allerdings auch zu bedenken, dass jeder Typ von Prüfmethode seine jeweiligen Schwächen und Stärken hat.

Brajnik (2008a) fasst verschiedene Typen von Prüfmethoden für die Barrierefreiheit im Web zusammen und diskutiert deren Vor- und Nachteile anhand bereits veröffentlichter wissenschaftlicher Arbeiten. Eine Art von manuellen Prüfmethoden, die von Brajnik genauer betrachtet wird, ist die bereits erwähnte Konformitätsprüfung, auch Expertenreview genannt. Als Vorteil zählt der Autor unter anderem auf, dass die Konformitätsprüfung eine große Bandbreite an verschiedenen Problemen für eine diverse Gruppe von Menschen mit Behinderung abdeckt. Die Qualität der Prüfung ist allerdings stark abhängig von dem zugrunde gelegten Standard. Beispielsweise können die WCAG als Standard dienen. Ein Vergleich der verschiedenen Prüfmethoden anhand konkreter Prüfverfahren bzw. Prüftools wird nicht durchgeführt. Brajnik merkt am Ende der Arbeit an, dass weitere Forschung notwendig ist, um zusätzliche Belege für die Vor- und Nachteile der jeweiligen Prüfmethoden zu identifizieren.

Darüber hinaus wurde untersucht, wie man manuelle und automatische Prüfmethoden kombinieren kann, um eine möglichst effektive Prüfmethode zu finden. Brajnik und Lomuscio (2007) haben ein neues Prüfverfahren entwickelt, das eine Kombination aus Konformitätsprüfung und automatischer Prüfung ermöglichen kann. Dieses Prüfverfahren wird jedoch im Rahmen der Arbeit weder mit anderen Prüfverfahren verglichen noch anhand von definierten Kriterien hinsichtlich seiner Qualität bewertet.

Graap et al. (2018) fassen diverse Prüftools und Prüfverfahren zur Bewertung der Barrierefreiheit von Websites zusammen. Dabei werden vier konkrete manuelle Prüfmethoden ausgewählt und genauer untersucht. Obwohl die verschiedenen Prüfverfahren von Graap et al. zur Prüfung von zehn Webseiten verwendet werden, werden lediglich Angaben über grundlegende Eigenschaften der Prüfverfahren gemacht. Nur vereinzelt erwähnen die Autoren die während der Prüfungen aufgetretenen Besonderheiten. Ein detaillierter und systematischer Vergleich der manuellen Prüfmethoden wird nicht durchgeführt. Des Weiteren werden zwei Prüfverfahren zur Zertifizierung barrierefreier Webseiten zusammengefasst. Die BIENE-Auszeichnung und der BIK BITV-Test, welcher auch in dieser Arbeit eine zentrale Rolle einnimmt. Eine Bewertung dieser beiden Prüfverfahren wird nicht vorgenommen.

Abschließend wird noch Bezug zu einer Studie von Baazeem und Al-Khalifa (2015) genommen, die sich mit den Fortschritten im Bereich der Prüfmethoden der Barrierefreiheit im Web in den Jahren 2011 bis 2015 auseinandersetzt. Basierend auf den gesammelten Daten kann gezeigt werden, dass immer die gleichen Prüfmethoden über diesen Zeitraum hinweg

verwendet wurden: automatische Prüfungen, manuelle Prüfungen durch Experten, Benutzertests mit Menschen mit Behinderung sowie Kombinationen dieser Prüfmethoden. Die Autoren kommen zu dem Fazit, dass ein Mangel an signifikanter Entwicklung der Prüfmethoden besteht und es einen großen Bedarf an weiteren Untersuchungen und Entwicklung der verfügbaren Methoden gibt.

4.3 Einflussfaktoren der manuellen Prüfmethoden

Die manuellen Prüfmethoden spielen bei der Bewertung der Barrierefreiheit von Websites eine wichtige Rolle. Daher haben sich Yesilada, Brajnik und Harper (2009) zum Ziel gesetzt, den Einfluss von Expertise auf die Bewertung der Barrierefreiheit zu untersuchen. Hierfür wurde eine Studie mit Experten und wenig erfahrenen Prüfern im Bereich Barrierefreiheit durchgeführt. Dabei müssen die Prüfer nicht nur die Barrieren identifizieren, sondern auch in ihrer Schwere bewerten. Es zeigt sich, dass der Grad der Expertise ein wichtiger Faktor für die Qualität der Bewertung ist. Zudem war die Effektivität sowie die Zuverlässigkeit der Experten höher als die der wenig erfahrenen Prüfer.

Ein weiterer wichtiger Einflussfaktor ist die Auswahl der zu prüfenden Seiten einer Website. Dabei ist zu beachten, dass eine geringere Anzahl an zu prüfenden Webseiten die benötigte Zeit der Prüfung und folglich auch die damit verbundenen Kosten reduziert. Eine qualitativ gute Webseitenauswahl soll den Grad der Barrierefreiheit der ganzen Website repräsentieren. Die Erstellung einer solchen repräsentativen Webseitenauswahl ist zentral für die Aussagekraft der gesamten Prüfung. Es wurden diverse Studien durchgeführt, um den Einfluss der Auswahlmethode und der Stichprobengröße zu untersuchen (vgl. Velleman und van der Geest, 2013; Brajnik, Mulas und Pitton, 2007). Brajnik et al. (2007) konnten in ihrer Untersuchung zeigen, dass die Genauigkeit stark von der verwendeten Auswahlmethode sowie der Stichprobengröße abhängt. Mit Genauigkeit meinen die Autoren den prozentualen Anteil der Verstöße einer Website bezüglich eines zugrunde gelegten Standards, der durch die Prüfung einer Webseitenauswahl erkannt werden kann. Eine Korrelation der Genauigkeit und der Struktur einer Website konnte nicht beobachtet werden. Zudem haben Zhang et al. (2015) eine spezielle Auswahlmethode entwickelt, die für Websites optimiert ist. Die Effektivität der Auswahlmethode für die Bewertung der Barrierefreiheit im Web wird durch experimentelle Ergebnisse gezeigt.

Eine häufig eingesetzte Art von manuellen Prüfmethoden ist die konformitätsbasierte Prüfmethode. Sie gründet auf anerkannten Standards. Diese sind maßgeblich für die Qualität

der Prüfung verantwortlich. Ein international anerkannter und häufig verwendeter Standard sind die bereits beschriebenen WCAG. Diverse Studien haben sich mit der Validität und Reliabilität der WCAG 1.0 sowie 2.0 befasst (vgl. Colwell und Petrie, 2001; Rømen und Svanæs, 2008; Brajnik, 2009; Calvo, Seyedarabi und Savva, 2016; Freire, 2012). Die Validität und Reliabilität spielen bei der Beurteilung der Qualität von Prüfmethoden eine wichtige Rolle. Rømen und Svanæs (2008) konnten in einer kontrollierten Usability-Studie mit Menschen mit Behinderung zeigen, dass nur 27% der gefundenen Barrierefreiheitsprobleme durch das Nutzen der WCAG 1.0 identifiziert werden konnten. Die Autoren kommen zu der Schlussfolgerung, dass die Anwendung der WCAG 1.0 allein nicht ausreicht, um die Barrierefreiheit einer Website zu gewährleisten. Ähnliche Ergebnisse werden bei der Studie von Calvo et al. (2016) für die WCAG 2.0 gefunden. Des Weiteren zeigt sich in einer Studie von Freire (2012), dass viele der durch beeinträchtigte Benutzer gefundenen Probleme nicht durch die Richtlinien (WCAG 1.0 und WCAG 2.0) abgedeckt werden. Zudem zeigt eine 2009 durchgeführte Studie von Brajnik, dass die WCAG hinsichtlich der Reliabilität im Allgemeinen schlecht abschneidet. In diesem Punkt stellt auch die WCAG 2.0 keine Verbesserung gegenüber der 1.0 dar.

4.4 Vergleich konkreter Prüfverfahren

Brajnik (2005) behauptet, dass das Fehlen von angemessenen Prüfmethoden einer der Gründe ist, warum Barrierefreiheit im Web derzeit schwer umzusetzen ist. Es werden diverse Arten von manuellen Prüfmethoden zur Bewertung der Barrierefreiheit kurz zusammengefasst, zum Beispiel die Konformitätsprüfung. Darüber hinaus wird ein neues Konzept für ein Prüfverfahren, das auf einem heuristischen Walkthrough[2] basiert, vorgestellt, aber nicht evaluiert. Der Autor schließt seine Arbeit mit der Forderung ab, die Prüfmethoden hinsichtlich ihrer Validität, Reliabilität sowie der Effektivität genau zu untersuchen. In einer späteren Arbeit erforscht Brajnik (2006) die von ihm bereits vorgestellte heuristische Walkthrough-Methode, im Folgenden Barrier-Walkthrough-Methode genannt. Für die Evaluation der Prüfmethoden identifiziert Brajnik folgende Kriterien: Validität, Reliabilität, Nützlichkeit und Effizienz. Bei der Wahl der Kriterien orientiert sich der Autor an früheren Arbeiten zu Evaluationsmethoden im Bereich der Usability. Die Ergebnisse der Studie zeigen, dass die Barrier-Walkthrough-Methode im Vergleich zur Konformitätsprüfung (genutzt wurde ein italienischer Standard) höhere Werte für die Validität und Nützlichkeit erzielt. Jedoch sind weitere Untersuchungen erforderlich, um die

[2]Dabei handelt es sich um eine Methode zur Evaluation der Usability einer Website (vgl. *Heuristischer Walkthrough*, 2018).

gefundenen Ergebnisse zu stützen, da es sich um ein Experiment von geringem Umfang handelt.

Brajnik (2008b) definiert die Vergleichskriterien noch weiter und testet diese anhand einer Studie. Es werden folgende Kriterien für den Vergleich von Prüfmethoden festgelegt: Wirksamkeit (Validität [Korrektheit, Empfindlichkeit], Reliabilität), Effizienz und Nützlichkeit. Er gibt einen kurzen Überblick über diverse Prüfmethoden sowie ihre Vor- und Nachteile. Anschließend wird der Vergleich anhand der vorher definierten Kriterien (ausgenommen Nützlichkeit) und zweier Prüfverfahren durchgeführt: der Konformitätsprüfung (genutzt wurde ein italienischer Standard) und der Barrier-Walkthrough-Methode. Der Vergleich zielt darauf ab, die Vorzüge der Barrier-Walkthrough-Methode zu bestimmen. Die Konformitätsprüfung wird daher als Kontrollbedingung verwendet. Die Ergebnisse zeigen Unterschiede beider Prüfmethoden bei der Validität sowie der Reliabilität. Jedoch wurden die Prüfungen von wenig erfahrenen Prüfern durchgeführt und könnten für erfahrene Prüfer andere Ergebnisse liefern.

Der BIK BITV-Test als Vertreter der konformitätsbasierten Prüfmethoden wird in zahlreichen wissenschaftlichen Studien als Metrik verwendet, um die Barrierefreiheit von Webangeboten zu beurteilen. Baumann (2012) untersucht mit diesem Prüfverfahren das Onlineangebot von Facebook auf Barrierefreiheit. Die Ergebnisse entsprechen alle dem Bewertungskriterium „schlecht zugänglich“. Er merkt an, dass die Barrierefreiheit nicht nur auf den Messwert des BIK BITV-Tests reduziert werden kann, aber dadurch zumindest ein Anhaltspunkt geliefert wird, wie zugänglich die Internetpräsenz ist. In einer Studie von Viehweger (2009) wird untersucht, in wie weit die Kriterien der Barrierefreiheit bei den Onlineangeboten von öffentlichen Bibliotheken umgesetzt sind. Zusätzlich zu dem BIK BITV-Test wird eine Usability-Studie mit Menschen mit Behinderung durchgeführt. Die Ergebnisse der beiden Prüfverfahren kommen zu sehr ähnlichen Ergebnissen. Alle drei geprüften Websites von öffentlichen Bibliotheken erfüllen nicht die Anforderungen der BITV. Es wird allerdings kein Vergleich der Prüfverfahren angestrebt. Viel mehr sollten die unterschiedlichen Prüfverfahren die Validität der Ergebnisse stützen.

Die existierenden Arbeiten untersuchen die Prüfmethoden mehr auf einer konzeptionellen Ebene oder stellen eigene Methoden vor. Ein umfassender systematischer Vergleich anhand von realen Daten, die durch Barrierefreiheitsexperten erstellt werden, fehlt. Um diesen systematischen quantitativen Vergleich im Rahmen dieser Arbeit durchführen zu können, wird ein generischer Kriterienkatalog erstellt. Er kann in Zukunft beliebig angepasst bzw. erweitert werden, um möglichen zukünftigen Anforderungen gerecht zu werden.

Mithilfe dieses erstellten Kriterienkatalogs werden zwei konkrete Prüfverfahren miteinander hinsichtlich ihrer Eignung und Wirksamkeit verglichen. Darunter das in Deutschland bekannteste Prüfverfahren: der BIK BITV-Test. Der Leser kann bei der Entscheidungsfindung, welches Prüfverfahren in der persönlichen Situation geeigneter ist, unterstützt werden. Zudem werden die verschiedenen Vor- und Nachteile detailliert aufgezeigt und es wird deutlich, welche Aspekte von welchem Prüfverfahren besser adressiert werden. Außerdem kann auch in dieser Arbeit beobachtet werden, dass die WCAG 2.1 nicht alle Probleme als solche ansehen, die jedoch laut Experteneinschätzungen für reale Benutzer bestehen. Damit erweitert diese Arbeit deutlich den Forschungsstand in diesem Bereich.

5 Bestehende Tools zur automatischen Prüfung

Im Folgenden wird eine Übersicht über derzeit verfügbare Prüftools zur Bewertung der Barrierefreiheit von Websites gegeben. Alle aufgeführten Prüftools wurden bereits in diversen wissenschaftlichen Arbeiten näher betrachtet (vgl. Burkard et al., 2021; Schmidt, 2013; Mateen, Voß-Nakkour und Rustemeier, 2020; Abduganiev, 2017). Die Übersicht hat jedoch nicht den Anspruch der Vollständigkeit, da es derzeit eine außerordentlich große Auswahl an verfügbaren Prüftools gibt. Zudem verändert sich der Bestand kontinuierlich: viele Tools werden weiterentwickelt, einige stehen jedoch auch nicht mehr zur Verfügung. Ein prominentes Beispiel hierfür ist der Onlinedienst AChecker, der seit dem 30.04.2021 nicht mehr erreichbar ist.

Abascal, Arrue und Valencia (2019) definieren Kriterien zur Klassifizierung von automatischen Prüfmethoden. Die folgende Auswahl der Kriterien wurde verwendet, um die Prüftools einander gegenüberzustellen: Nutzung (kostenlos vs. kostenpflichtig), Plattform (lokale Anwendung vs. Onlinedienst), Evaluationsumfang (einzelne Webseite, Auswahl an Webseiten vs. ganze Website) und verwendete Standards. Zudem wurden die Prüftools in diese drei Kategorien aufgeteilt: Validierungswerkzeuge, Werkzeuge zur Prüfung von Farben sowie Kontrasten und WCAG-/BITV-Testwerkzeuge (vgl. Schmidt, 2013).

Name	Nutzung	Plattform
HTML Tidy[1]	kostenlos	lokale Anwendung
W3C Markup Validation Service[2]	kostenlos	Onlinedienst
W3C CSS Validation Service[3]	kostenlos	Onlinedienst

Tabelle 5.1: Validierungswerkzeuge

Name	Nutzung	Plattform
Colorblind Web Page Filter[4]	kostenlos	Onlinedienst
WCAG Contrast Checker[5]	kostenlos	Onlinedienst

Tabelle 5.2: Werkzeuge zur Prüfung von Farben und Kontrasten

Validierungswerkzeuge untersuchen den Quellcode bzw. die Markup-Sprache hinsichtlich ihrer Korrektheit. Außerdem können einige Prüftools auch einfachere Korrekturen direkt vornehmen. Tabelle 5.1 zeigt drei Validierungswerkzeuge.

Zwei Werkzeuge zur Prüfung von Farben und Kontrasten sind in Tabelle 5.2 aufgelistet. Mithilfe dieser Prüftools können Farben und Kontraste im Kontext von Barrierefreiheit untersucht werden. Andere Aspekte der Barrierefreiheit werden von diesen Prüftools außer Acht gelassen.

WCAG-/BITV-Testwerkzeuge prüfen Webseiten anhand von Standards wie WCAG oder der BITV. Tabelle 5.3 beinhaltet zehn solcher Prüftools. Viele der hier aufgeführten Prüftools sind in verschiedenen Versionen erhältlich und stehen in den meisten Fällen auch als Browsererweiterungen oder Bookmarklets zur Verfügung.

Mithilfe der in diesem Kapitel zusammengefassten Prüftools kann ein erster Eindruck der Barrierefreiheit einer Website gewonnen werden. Die Prüftools können allerdings nicht feststellen, ob eine Website tatsächlich barrierefrei ist. Derzeit können nur manuelle Prüfmethoden eine umfassende Prüfung auf Barrierefreiheit gewährleisten. Jedoch können Prüftools als Hilfsmittel bei einer manuellen Prüfung eingesetzt werden. Beispielsweise wird eine Prüfung mit dem W3C Markup Validation Service standardmäßig im Rahmen des BIK BITV-Test durchgeführt.

[1] von HTML Tidy Advocacy Community Group verfügbar unter `http://www.html-tidy.org/`
[2] von W3C verfügbar unter `https://validator.w3.org/`
[3] von W3C verfügbar unter `https://jigsaw.w3.org/css-validator/`
[4] von Toptal verfügbar unter `https://www.toptal.com/designers/colorfilter`
[5] von WebAIM verfügbar unter `https://webaim.org/resources/contrastchecker/`

Name	Nutzung	Plattform	Umfang	Standard
Siteimprove Accessibility Checker[7]	kostenlos	Browser-erweiterung	einzelne Webseite	WCAG 2.0
WAVE Browser Extension[8]	kostenlos	Browser-erweiterung	einzelne Webseite	WCAG 2.1
CynthiaSays[9]	kostenlos	Online-dienst	einzelne Webseite	WCAG 2.0
WAVE[10]	kostenlos	Online-dienst	einzelne Webseite	WCAG 2.1
Mauve[11]	kostenlos	Online-dienst	Auswahl an Webseiten	WCAG 2.1
Total Validator (Pro Version)[12]	kostenpflichtig	lokale Anwendung	ganze Website	WCAG 2.1
axe Monitor[13]	kostenpflichtig	Online-dienst	ganze Website	WCAG 2.1
Accessibility[14]	kostenpflichtig	Online-dienst	ganze Website	WCAG 2.1
ARC Monitoring (Essentials Version)[15]	kostenpflichtig	Online-dienst	ganze Website	WCAG 2.1
Pope Tech[16]	kostenpflichtig	Online-dienst	ganze Website	WCAG 2.1

Tabelle 5.3: WCAG-/BITV-Testwerkzeuge

Eine umfangreiche Liste von Prüftools ist unter der Website Web Accessibility Evaluation Tools List[6] zu finden. Die Website wird von W3C seit 2016 zur Verfügung gestellt. Die dort gesammelten Informationen zu den Prüftools stammen jedoch von den verschiedenen Anbietern oder Dritten (vgl. W3C, 2016).

[6] verfügbar unter `https://www.w3.org/WAI/ER/tools/`
[7] von Siteimprove verfügbar unter `https://siteimprove.com/de-de/core-platform/integrations/browser-extensions/`
[8] von Webaim verfügbar unter `https://wave.webaim.org/extension/`
[9] von Compliance Sheriff®, ICDRI, und der Internet Society Disability and Special Needs Chapter verfügbar unter `http://www.cynthiasays.com/`
[10] von Webaim verfügbar unter `https://wave.webaim.org/`
[11] von HIIS Lab des Institute of Information Science and Technologies von National Research Council verfügbar unter `https://mauve.isti.cnr.it/index.jsp`
[12] von Total Validator verfügbar unter `https://www.totalvalidator.com/`
[13] von Deque verfügbar unter `https://www.deque.com/axe/monitor/`
[14] von Siteimprove verfügbar unter `https://siteimprove.com/de-de/barrierefreiheit/`
[15] von The Paciello Group verfügbar unter `https://www.tpgi.com/arc-platform/monitoring/`
[16] von PopeTech verfügbar unter `https://pope.tech/`

6 Bestehende Verfahren zur manuellen Prüfung

Im Folgenden wird eine Übersicht der derzeit bestehenden Prüfverfahren im Bereich der manuellen Prüfmethoden zur Bewertung der Barrierefreiheit gegeben. Die Liste hat keinen Anspruch auf Vollständigkeit. Es hat sich bei der Recherche gezeigt, dass immer wieder neue Prüfverfahren entwickelt und bestehende aktualisiert werden. Manche Prüfverfahren jedoch verlieren an Relevanz, wenn sie nicht weiterentwickelt werden. Ein Beispiel hierfür sind die Kriterien und Prüfschritte des BIENE Awards[1]. Auch ein in der Vergangenheit bekanntes Prüfverfahren zur Selbstbewertung der Barrierefreiheit, Banu, ist nicht in der Übersicht enthalten, da auf der offiziellen Internetpräsenz[2] seit Monaten die Meldung „Diese Seite wird derzeit überarbeitet“ angezeigt wird.

Die Kriterien zur Klassifizierung der manuellen Prüfmethoden wurden in Anlehnung an die Kriterien der automatischen Prüfmethoden gewählt und um folgende wichtige Kriterien ergänzt: Qualifikation des Prüfers (Selbstbewertung vs. Expertenbewertung), Transparenz des Prüfverfahrens (öffentlich vs. nicht öffentlich), Anpassbarkeit des zugrunde gelegten Standards und Zertifizierung (möglich vs. nicht möglich). Die Transparenz des Prüfverfahrens gibt an, ob eine detaillierte Vorgehensbeschreibung inklusive der einzelnen durchzuführenden Prüfschritte öffentlich verfügbar ist oder nicht. Die Prüfverfahren werden anhand der Qualifikation des Prüfers in zwei Arten aufgeteilt: Experten- und Selbstbewertungen.

[1] Von 2013 bis 2010 wurde eine Auszeichnung für die besten barrierefreien Webseiten von der Aktion Mensch und der Stiftung Digitale Chancen vergeben (vgl. Aktion Mensch und Stiftung Digitale Chancen, o.D.).

[2] verfügbar unter `https://www.banu.bund.de`

Name	Transparenz	Standard	Anpass-barkeit	Zertifizierung
BIK BITV-Test (vor März 2021)[3]	öffentlich	WCAG 2.1 AA	nein	möglich
BIK BITV-Test (ab März 2021)[4]	öffentlich	WCAG 2.1 AA, EN 301 549	nein	möglich
BITV-Audit[5]	nicht öffentlich	WCAG 2.1 AAA, EN 301 549 BITV 2.0	ja	möglich
BIKOSAX[6]	nicht öffentlich	WCAG 2.1, EN 301 549	nein	möglich
Barriere-Check Pro[7]	nicht öffentlich	WCAG 2.1, BITV 2.0, Aspekte der Usability und Performance	ja	nicht möglich

Tabelle 6.1: Übersicht der Prüfverfahren zur Expertenbewertung

6.1 Expertenbewertung

Tabelle 6.1 zeigt die Prüfverfahren, die der Kategorie Expertenbewertungen zugeordnet werden. Diese Prüfungen werden durch einen qualifizierten Prüfer vollzogen. Alle aufgeführten Prüfverfahren führen die Prüfung anhand einer repräsentativen Webseitenauswahl durch, um so die Barrierefreiheit der ganzen Website widerspiegeln zu können. Folgende Prüfverfahren haben ihr detailliertes Vorgehen nicht veröffentlicht: BITV-Audit, BIKOSAX und Barriere-Check Pro. Das bedeutet, dass die einzelnen Prüfschritte nicht öffentlich einsehbar sind, sondern lediglich die zugrunde gelegten Standards definiert werden. Der BITV-Audit und der Barriere-Check Pro können je nach Kundenwunsch das Prüfvorgehen anpassen, um so einer Abstufung des maximal erreichbaren Standards zu entsprechen. Beispielsweise kann der BITV-Audit adaptiert werden und so nur WCAG 2.1 AA prüfen. Bis auf das Prüfverfahren Barriere-Check Pro können alle die Barrierefreiheit einer Website bei erfüllter Konformität zertifizieren. Der BIK BITV-Test kommt in der Übersicht zweimal vor, da bei einer Aktualisierung im März 2021 weitreichende Än-

[3] von BIK
[4] von BIK verfügbar unter `https://www.bitvtest.de/start.html`
[5] von T-Systems MMS
[6] von dzb lesen verfügbar unter `https://www.dzblesen.de/index.php?site_id=4.5`
[7] von anatom5 perception marketing GmbH verfügbar unter `https://barrierekompass.de/leistungen/barriere-check-pro.html`

derungen am Prüfverfahren vorgenommen wurden. Daher werden diese beiden Versionen differenziert voneinander betrachtet.

6.2 Selbstbewertung

Die Prüfverfahren zur Selbstbewertung der Barrierefreiheit sind in Tabelle 6.2 zu finden. Sie sind alle kostenlos und ihr genaues Prüfvorgehen ist veröffentlicht. Jedoch kann keins der Prüfverfahren zu einer Zertifizierung der Barrierefreiheit genutzt werden. Folgende vier Prüfverfahren unterstützen den Prüfer durch einen Onlinedienst bei der Prüfungsdurchführung: die BITV/WCAG Selbstbewertung vor März sowie ab März 2021, die Schweizer-Accessibility Checkliste 2.1 und WCAG EM. Einige der Prüfverfahren bieten darüber hinaus Optionen zur Erstellung eines Prüfberichts an. Bis auf das Prüfverfahren WCAG EM umfassen alle anderen bei der Prüfung nur eine einzelne Webseite und keine repräsentative Webseitenauswahl. Die Prüfverfahren geben keine Entscheidungshilfe, welche Webseiten bei der Auswahl eingeschlossen werden sollen. Ebenso sind die Onlinedienste nicht zur Dokumentation und Bewertung der gefundenen Barrieren zu mehreren Webseiten ausgelegt. Infolgedessen kann mithilfe der Ergebnisse der Prüfverfahren nur eine Aussage zur Barrierefreiheit einer einzelnen Webseite getroffen werden, nicht über die ganze Website. In den zugrunde liegenden Standards variieren die Prüfverfahren genauso wie bei der Anpassbarkeit des Prüfvorgehens. Folgende fünf Prüfverfahren sehen keine Adaption der Standards vor: BITV/WCAG Selbstbewertung vor sowie nach März 2021, Quick Check, Easy Checks und die Checkliste für Usability Professionals. Der BIK BITV-Test kommt auch in dieser Übersicht zweimal vor, da bei der Aktualisierung im März 2021 ebenfalls die Selbstbewertung überarbeitet wurde.

Name	**Unterstützung bei Prüfung**	**Umfang**	**Standard**	**Anpass-barkeit**
BITV/WCAG Selbstbewertung (vor März 2021)[8]	Online-dienst	einzelne Webseite	WCAG 2.1 AA	nein
BITV/WCAG Selbstbewertung (ab März 2021)[9]	Online-dienst	einzelne Webseite	WCAG 2.1 AA EN 301 549	nein
Schweizer Accessibility-Checkliste 2.1[10]	Online-dienst	einzelne Webseite	WCAG 2.1 AA	ja
WCAG EM (+ Report Tool)[11]	Online-dienst	repräsent. Webseiten-auswahl	WCAG 2.1 AAA	ja
IBM Accessibility Requirements[12]	keine Tool-unterstützung	einzelne Webseite	WCAG 2.1 AA EN 301 549	ja
Quick Check[13]	keine Tool-unterstützung	einzelne Webseite	einzelne Prüf-schritte des BIK BITV-Test	nein
Easy Checks[14]	keine Tool-unterstützung	einzelne Webseite	einzelne Erfolgs-kriterien der WCAG 2.1	nein
Checkliste für Usability Professionals[15]	keine Tool-unterstützung	einzelne Webseite	einzelne Prüf-schritte des BIK BITV-Test	nein

Tabelle 6.2: Übersicht der Prüfverfahren zur Selbstbewertung

[8] von BIK verfügbar unter https://testen.bitv-test.de/selbstbewertung/
[9] von BIK verfügbar unter https://webtest.bitv-test.de/selbstbewertung/
[10] von Zugang für alle verfügbar unter https://www.access-for-all.ch/ch/barrierefreiheit/barrierefreies-webdesign/accessibility-checkliste-2-1.html
[11] von WAI (W3C) verfügbar unter https://www.w3.org/TR/WCAG-EM/
[12] von IBM verfügbar unter https://www.ibm.com/able/requirements/overview
[13] von anatom5 perceotion marketing GmbH verfügbar unter https://barrierekompass.de/barrierefreies-internet/8-einfache-tests-testen-sie-selbst.html
[14] von WAI (W3C) verfügbar unter https://www.w3.org/WAI/test-evaluate/preliminary/
[15] von German UPA verfügbar unter https://www.germanupa.de/sites/default/files/public/content/2018/2018-02-10/webchecklistecd3germanupabfpdfua.pdf

6.3 Auswahl der Prüfverfahren für den Vergleich

Um die Unterschiede, Gemeinsamkeiten, Vor- und Nachteile der beiden Arten (konformitätsbasiert vs. empirisch ermittelt) von manuellen Prüfmethoden gut identifizieren zu können, muss jeweils ein Vertreter ausgewählt werden. Zur Durchführung eines expertenbasierten Vergleichs kommen nur Prüfverfahren in Frage, deren Vorgehen den Einsatz von Experten vorsehen. Zudem sollten beide Vertreter die Möglichkeit bieten, die Barrierefreiheit einer Website bei erfüllter Konformität zu zertifizieren, um vergleichbare Prüfungsergebnisse zu liefern. Für den Vergleich werden folgende Prüfverfahren ausgewählt: der BIK BITV-Test (konformitätsbasiert) und der BITV-Audit der T-Systems MMS (empirisch ermittelt). Der BIK BITV-Test ist das in Deutschland etablierte Prüfverfahren zur Konformitätsbeurteilung von Websites und wurde aus diesem Grund ausgewählt. Im Rahmen dieser Arbeit wird ausschließlich die Version vor März 2021 betrachtet. Es wird keine Aussage zur aktuellen Version des Prüfverfahrens gemacht. Als Vertreter der empirisch ermittelten Prüfverfahren wurde der BITV-Audit ausgewählt. Das Prüfverfahren besteht aus einer Zugänglichkeitsprüfung und einer Konformitätsprüfung. Bei der empirischen Expertenevaluation wird die Zugänglichkeit aus der Perspektive von Menschen mit Behinderung betrachtet. Im Anschluss werden die gefundenen Probleme den Konformitätskriterien zugeordnet und überprüft, ob noch nicht berücksichtigte Kriterien verletzt wurden. Daher ist auch durch dieses Prüfverfahren eine Zertifizierung von Websites möglich.

7 Detailbeschreibung der ausgewählten Prüfverfahren

Im Folgenden wird dem Leser ein Überblick über die ausgewählten Prüfverfahren gegeben, um so den späteren Vergleich nachvollziehbar zu machen. Abschnitt 7.1 gibt einen detaillierten Einblick in das Prüfvorgehen des BIK BITV-Tests. Dabei werden die Besonderheiten des Prüfverfahrens herausgearbeitet. Im Abschnitt 7.2 wird anschließend auf das Prüfverfahren BITV-Audit von T-Systems MMS und dessen Besonderheiten eingegangen.

7.1 BIK BITV-Test

Beim BIK BITV-Test handelt es sich um ein Prüfverfahren zur Bewertung der Barrierefreiheit von Websites und Webanwendungen. Das Prüfverfahren wurde vom BIK-Projekt zusammen mit Barrierefreiheitsexperten, Behindertenverbänden und Dienstleistern im Bereich Web erarbeitet und ist vollständig offengelegt. Die Erstveröffentlichung erfolgte im Jahr 2004 und das Prüfverfahren wird seitdem kontinuierlich weiterentwickelt (vgl. DIAS GmbH, 2020d). Die letzte Überarbeitung wurde erst im März 2021 veröffentlicht (vgl. DIAS GmbH, 2021). In den nachfolgenden Kapiteln wird jedoch ausschließlich Bezug auf die Version des Prüfverfahrens vor März 2021 genommen.

7.1.1 Zugrunde gelegter Standard

Dem Prüfverfahren zugrunde gelegt sind folgende Standards: BITV 2.0, EN 301 549 (Abschnitt 9 Web) und WCAG 2.1 AA. Die zusätzlichen Anforderungen der BITV 2.0 werden nicht im Prüfverfahren abgedeckt. Zudem sind auch die zusätzlichen Anforderungen der Tabelle A.1 der EN 301 549 in der Version vor März 2021 noch nicht aufgenommen. Es

gibt außerdem noch keine Prüfschritte, die die WCAG 2.1 Erfolgskriterien mit Konformitätsstufe AAA umsetzen. Das Prüfen von Nicht-Web-Dokumenten (z.B. PDF-Dokumente) ist nicht Teil des Prüfverfahrens (vgl. DIAS GmbH, 2019).

7.1.2 Varianten des Prüfverfahrens

Es gibt drei Varianten des BIK BITV-Tests: Entwicklungsbegleitender BITV-Test, abschließender BITV-Test und BITV-Selbstbewertung. Der entwicklungsbegleitende BITV-Test bietet sich, wie der Name schon andeutet, besonders während der Entwicklung neuer Websites und Webanwendungen zur Qualitätssicherung der Barrierefreiheit an. Der abschließende BITV-Test kann eingesetzt werden, um den erreichten Grad an Barrierefreiheit eines Webangebots zu ermitteln. Die BITV-Selbstbewertung kann während der Entwicklung ohne Hinzuziehen von Experten verwendet werden, wenn die Vergleichbarkeit der Ergebnisse nicht im Vordergrund steht. Alle drei Varianten haben den gleichen Umfang an Prüfschritten, unterscheiden sich jedoch beim Prüfvorgehen. Während eine Zertifizierung der Konformität nur durch einen abschließenden BITV-Test erreicht werden kann, dürfen die Ergebnisse des entwicklungsbegleitenden BITV-Tests und der BITV-Selbstbewertung nur für den internen Gebrauch verwendet werden (vgl. DIAS GmbH, 2020e). Im weiteren Verlauf dieser Arbeit wird ausschließlich Bezug auf den abschließenden BITV-Test genommen.

7.1.3 Prüfvorgehen

Abbildung 7.1 zeigt das Prüfvorgehen. Zu Beginn wird zusammen mit dem Auftraggeber die passende Prüfvariante ausgewählt, je nachdem welches Ziel mit der Prüfung verfolgt werden soll. Es wird geklärt welche Bereiche und Webseiten zum Webangebot gehören, welches mit dem BITV-Test auf Konformität geprüft werden sollen. Anschließend kann der erforderliche Umfang und Preis geschätzt werden. Der Erstprüfer legt die Prüfdaten im BITV-Test-Prüftool an und wählt die zu prüfenden Webseiten aus. Die Prüfung der ganzen Website ist nicht umsetzbar, da eine solche Prüfung zu aufwendig wäre. Daher wird eine repräsentative Webseitenauswahl getroffen, diese Auswahl erfolgt ohne den Auftraggeber. Die Größe der Stichprobe ist abhängig von der Komplexität des Prüfgegenstands. Die Qualitätssicherung prüft die Webseitenauswahl und gibt sie zur Prüfung frei. Die Prüfung der einzelnen Prüfschritte wird von zwei Prüfern des BIK-Prüfverbunds durchgeführt. Erst nachdem jeder Prüfer seine Prüfung abgeschlossen hat, besprechen sie die

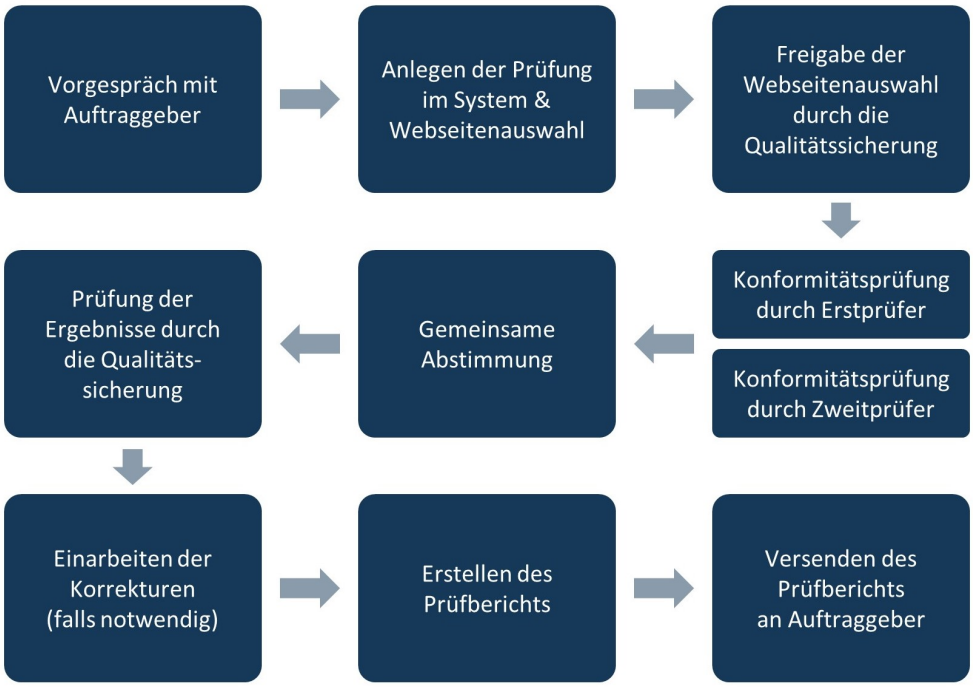


Abbildung 7.1: BIK BITV-Test: Schematische Darstellung des Prüfvorgehens

gefundenen Ergebnisse in einer gemeinsamen Abstimmung. Im Anschluss wird eine Qualitätssicherung der gefundenen Ergebnisse durch eine unabhängige qualifizierte Prüfstelle durchgeführt. Werden hier notwendige Korrekturen der Ergebnisse identifiziert, werden diese durch den Erstprüfer überarbeitet. Optional können hier mehrere Korrekturschleifen durchlaufen werden. Sobald diese Korrekturarbeiten abgeschlossen sind, kann die Qualitätssicherung den Prüfbericht im BITV-Test-Prüftool generieren lassen und der Erstprüfer versendet diesen an den Auftraggeber (vgl. DIAS GmbH, 2019). Der Prüfbericht kann veröffentlicht werden und als Zertifizierung der Barrierefreiheit auf der Website verlinkt werden.

7.1.4 Toolunterstützung

Das BITV-Test-Prüftool (DIAS GmbH, o.D.d) ist eine Webanwendung und unterstützt die Prüfer des BIK-Prüfverbunds bei der Prüfungsdurchführung und der Prüfberichterstellung. Ein neuer Prüfauftrag kann vom Erstprüfer in der Webanwendung angelegt werden. Die Benutzeroberfläche zum Anlegen eines neuen Prüfauftrags ist in Abbildung 7.2 zu sehen. Bei einem abschließenden BIK BITV-Test, weist der Erstprüfer die Rolle des Zweitprüfers einem weiteren Prüfer des BIK-Prüfverbunds zu. Zudem wählt er aus, wer

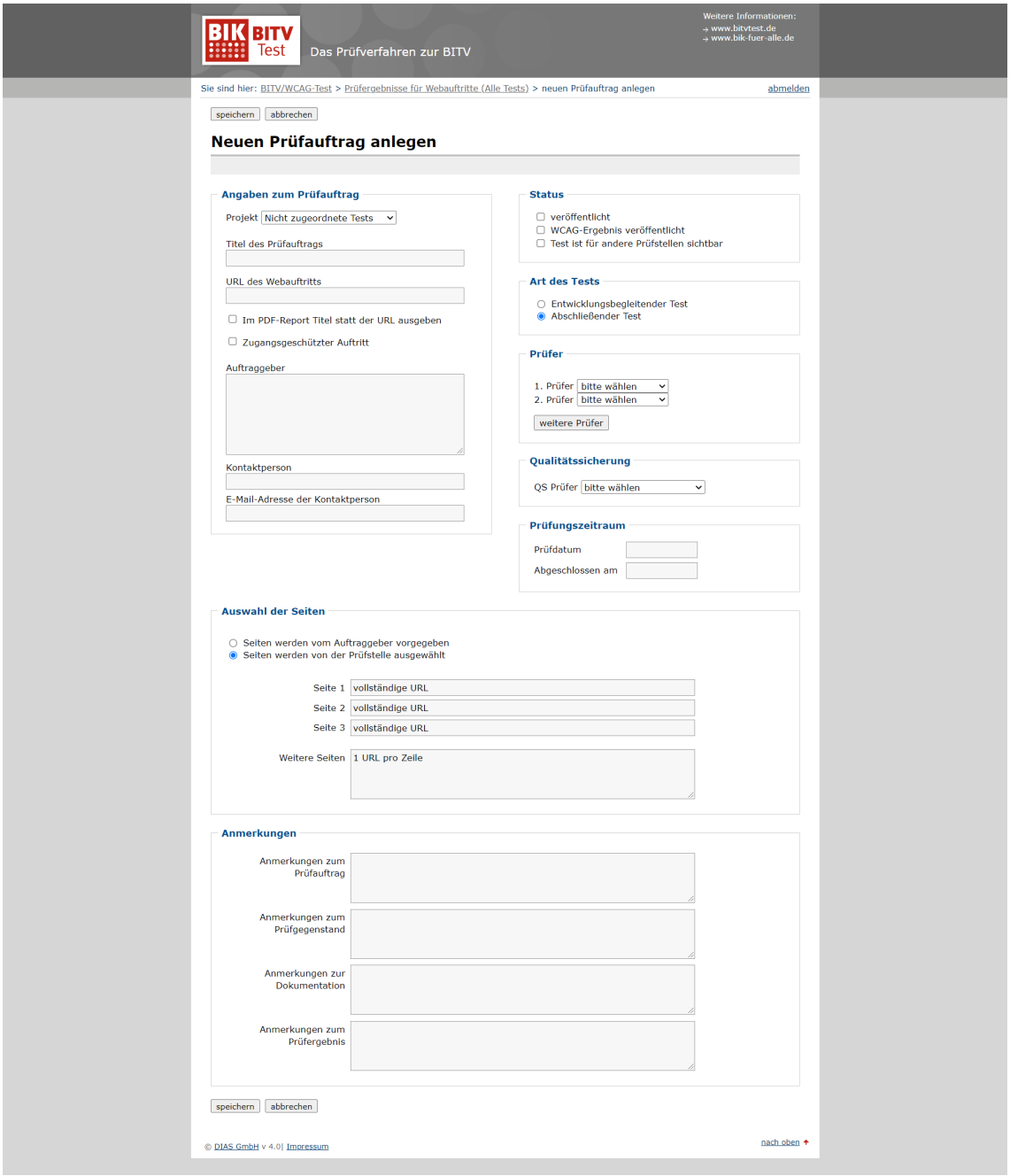


Abbildung 7.2: BIK BITV-Test: Benutzeroberfläche zum Anlegen eines neuen Prüfauftrags

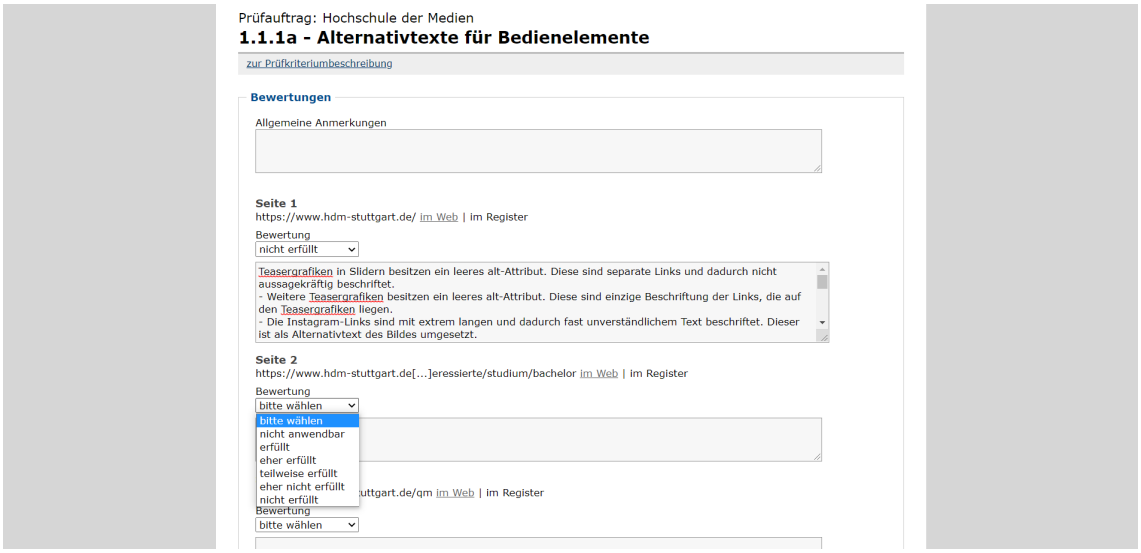


Abbildung 7.3: BIK BITV-Test: Ausschnitt der Benutzeroberfläche zur Prüfung des Prüfschritts 1.1.1a - Alternativtexte für Bedienelemente

die Qualitätssicherung übernimmt. Er gibt gleichzeitig auch die Webseitenauswahl an. Den beteiligten Prüfern stehen je nach Rolle (Qualitätssicherung, Erst- oder Zweitprüfer) unterschiedliche Funktionen bei dem Prüfauftrag zur Verfügung. Die Qualitätssicherung gibt die ausgewählten Webseiten zur Prüfung frei. Anschließend können die beiden Prüfer, unabhängig voneinander, nacheinander für jeden Prüfschritt eine Bewertung des Prüfschritts sowie eine Beschreibung der gefundenen Probleme für jede Webseite notieren. Abbildung 7.3 zeigt diese Benutzeroberfläche. Alle beteiligten Prüfer können jederzeit den Stand der Bearbeitung einsehen. Beispielsweise kann der Status der Prüfung des Erst- und Zweitprüfers auf „in Arbeit“, „abgeschlossen“ oder „abgebrochen“ stehen. Sobald beide Prüfer ihre Prüfung abgeschlossen haben, können sie sich für die gemeinsame Abstimmung verabreden und die Dokumentation der Probleme sowie Bewertung der Prüfschritte gemeinsam überarbeiten. Wenn die Abstimmung erfolgreich beendet wurde, kann die Qualitätssicherung die abgestimmten Ergebnisse einsehen und prüfen. Über notwendige Korrekturen informiert die Qualitätssicherung den Erstprüfer per E-Mail. Wenn etwaige Korrekturen im BITV-Test-Prüftool abgeschlossen sind, kann die Qualitätssicherung den PDF-Prüfbericht von der Anwendung generieren lassen. Der Erstprüfer kann dann den Prüfbericht an den Auftraggeber senden.

7.1.5 Prüferqualifikation

Die Durchführung des BIK BITV-Tests wird von erfahrenen Experten vorgenommen. Die Experten müssen zuvor ein Verfahren zur Qualifizierung durchlaufen, um offizieller BIK BITV-Prüfer werden zu können. Hierfür muss der Experte ein grundlegendes Verständnis von HTML besitzen, aber vor allem das BITV-Test Prüfverfahren kennen und die Prüfschritte sicher anwenden können. Einrichtungen mit qualifizierten Prüfern werden dann als BITV-Test-Prüfstelle in den BIK-Prüfverbund aufgenommen (vgl. DIAS GmbH, 2019).

7.1.6 Qualifizierung einer Organisation als Prüfstelle

Die DIAS GmbH ist die Ausbildungsstelle für den BIK BITV-Test, hat jedoch begrenzte Kapazitäten. Es gibt sehr viele Anfragen von Organisationen, die sich als Prüfstelle qualifizieren wollen. Allerdings schreckt auch die Anforderung, dass das eigene Webangebot barrierefrei sein muss, einige Interessenten ab. Die Interessenten kommen, bei grundsätzlicher Eignung, auf eine Warteliste, wobei nicht-kommerzielle Stellen (z.B. Selbsthilfeverbände) bevorzugt ausgebildet werden. Mindestens ein Mitarbeiter der Organisation muss sich als Prüfer qualifizieren. Im Rahmen der Qualifizierung müssen drei beauftrage BIK BITV-Tests als Zweitprüfer durchgeführt werden. Dafür muss sich der Anwärter die Zeit nehmen die Dokumentation der Prüfschritte zu verstehen und das Angebot anhand dieser Prüfschritte eigenverantwortlich zu prüfen. Im Anschluss erfolgt die gemeinsame Abstimmung mit einem erfahrenen Prüfer. Schließlich wird eine statistische Auswertung genutzt, um die Abweichungen zwischen den Ergebnissen des Anwärters und den finalen Bewertungen zu vergleichen. Dabei darf ein definierter Grenzwert nicht überschritten werden (D. Fischer [DIAS GmbH], persönliche Kommunikation, 05. März 2021).

7.1.7 Prüfschritte

Der BITV-Test umfasst unabhängig vom Prüfgegenstand immer insgesamt 60 Prüfschritte. Diese sind öffentlich zugänglich und definieren was genau geprüft wird, wieso das wichtig ist und wie das Vorgehen der einzelnen Prüfschritte aussieht (vgl. DIAS GmbH, o.D.f). Die Prüfschritte werden nicht gewichtet. Ein Großteil der Prüfschritte ist nur unter bestimmten Voraussetzungen anwendbar. Diese Voraussetzungen prüft der Barrierefreiheitsexperte bei jedem Prüfschritt im Laufe der Prüfung (vgl. DIAS GmbH, 2019).

7.1.8 Bewertungskategorien

Die Bewertungen der Prüfschritte können folgenden fünf Kategorien entsprechen: erfüllt, eher erfüllt, teilweise erfüllt, eher nicht erfüllt und nicht erfüllt. Wann welche Kategorie zu wählen ist, hängt von dem jeweiligen Prüfschritt ab und ist ebenfalls in der Prüfschrittbeschreibung dokumentiert.

Eine BITV-Anforderung gilt nur dann als „erfüllt“, wenn alle zugehörigen Prüfschritte „erfüllt“ oder „eher erfüllt“ sind. Sobald ein Prüfschritt eine andere Bewertung erhält, gilt die übergeordnete BITV-Anforderung als „nicht erfüllt“. Für die Konformitätsaussage gilt, dass sobald eine BITV-Anforderung „nicht erfüllt“ ist, die geprüfte Webseite nicht BITV-konform ist (vgl. DIAS GmbH, 2019).

7.1.9 Qualitätssicherung

Durch die unabhängige Prüfung der Webseitenauswahl durch zwei Prüfer und deren gemeinsame Abstimmung soll eine gleichbleibende Qualität gewährleistet und etwaige Bewertungsspielräume in ähnlicher Weise genutzt werden. In der Regel gehören die beiden Prüfer unterschiedlichen Prüfstellen an. Zusätzlich wird beim abschließenden BITV-Test eine Qualitätssicherung durch eine qualifizierte Prüfstelle durchgeführt. Dabei wird stichprobenartig geprüft, ob die Bewertungen korrekt sind. So kann eine einheitliche Bewertungspraxis sichergestellt werden. Der für die Qualitätssicherung verantwortliche Prüfer gibt dem Erst- und Zweitprüfer Rückmeldung, sodass diese die Prüfergebnisse nochmals korrigieren können. Diese Korrekturschleife kann optional wiederholt werden (vgl. DIAS GmbH, 2019).

7.1.10 Prüfbericht

Der PDF-Prüfbericht beginnt mit den formalen Angaben zur Prüfung. Dann gibt er einen Überblick über die BITV-Konformität und die WCAG-Konformität der geprüften Webseiten. Außerdem gibt es einen Abschnitt „Anmerkungen zu einzelnen Prüfschritten“. Hier werden alle Bewertungen der einzelnen Prüfschritte sowie die gefundenen Probleme der einzelnen Prüfschritte aufgelistet (vgl. DIAS GmbH, 2019).

Abbildung 7.4: BIK BITV-Test: BITV-Prüfzeichen (DIAS GmbH, o.D.g)

7.1.11 Veröffentlichung der Ergebnisse

Die Prüfberichte von abschließenden BIK BITV-Tests können durch den Auftraggeber veröffentlicht werden. Jedoch muss der Prüfbericht vollständig offen gelegt werden. Es dürfen keine Teile ausgelassen werden. Bei BITV-konformen Ergebnis kann das Webangebot auch auf der Website des BIK BITV-Tests in die Liste der barrierefreien Websites aufgenommen werden (vgl. DIAS GmbH, 2019).

7.1.12 Prüfzeichen

Abbildung 7.4 zeigt die beiden BITV-Prüfzeichen: BIK BITV-Test und BIK BITV-konform (geprüfte Seiten). Das Prüfzeichen „BIK BITV-Test“ gibt an, dass ein abschließender BIK BITV-Test durch eine offizielle Prüfstelle durchgeführt wurde. Es sagt allerdings nichts über das Ergebnis des Tests aus. Soll das Prüfzeichen „BIK BITV-Test“ innerhalb eines Webangebots verwendet werden, muss es auf die EzB des Webangebots verlinken und dort die gefundenen Mängel offenlegen. Alternativ kann auch auf den online Prüfbericht auf der BIK BITV-Test Website verlinkt werden. Das zweite Prüfzeichen „BIK BITV-konform“ jedoch steht für einen abschließenden BIK BITV-Test, dessen Ergebnisse zeigen, dass alle BITV-Anforderungen erfüllt werden. Auch dieses Prüfzeichen muss auf die EzB oder den online Prüfbericht verlinken (vgl. DIAS GmbH, 2019).

7.1.13 Preise

Der Preis setzt sich aus der Komplexität und dem Umfang des Prüfgegenstands zusammen. Hierfür werden vier Komplexitätsstufen unterschieden, siehe Tabelle 7.1. Der Gesamtpreis entspricht der Summe aller Preise für die einzelnen zu prüfenden Webseiten (vgl. DIAS GmbH, 2020a).

Komplexitätsstufe	Techniken (Beispiele)	Preis pro Seite
I	Einfaches HTML	558 €
II	Umfangreiche Seiten, Seiten mit dynamischen Inhaltselementen	837 €
III	Umfangreiche oder komplexe Seiten, Seiten mit verschiedenartigen dynamischen Inhaltselementen	1.116 €
IV	Seiten bzw. Ansichten von komplexen Webanwendungen, mehrstufige Prozesse, Seiten mit umfangreichen Inhalten oder vielen dynamischen Inhaltselementen	Preis ist abhängig vom kalkulierten Prüfaufwand

Tabelle 7.1: BIK BITV-Test: Preis pro Webseite für verschiedene Komplexitätsstufen (DIAS GmbH, 2020a)

7.1.14 Änderungen der Version ab März 2021

Im Rahmen der Überarbeitung im März 2021 wurden die zusätzlichen Anforderungen der EN 301 549, die in Tabelle A.1 definiert werden, in das Prüfverfahren aufgenommen. Daher erhöht sich die Anzahl der Prüfschritte von 60 auf 92 (vgl. DIAS GmbH, 2021). Zudem wurde das Tandemverfahren abgeschafft. Folglich wird die Prüfung nur noch von einem qualifizierten Prüfer durchgeführt. Die Qualitätssicherung nimmt nun eine noch gewichtigere Rolle ein. Außerdem gibt es keine Unterscheidung mehr zwischen einem entwicklungsbegleitenden und einem abschließenden BITV-Test. Es gibt allerdings die Möglichkeit einen Test für interne Zwecke zu beauftragen. In diesem Fall kann der Auftraggeber, die Webseitenauswahl vorgeben. Die Prüfergebnisse eines internen Tests dürfen jedoch nicht veröffentlicht werden. Im Zuge der Überarbeitung gab es auch eine Preisanpassung der Einzelpreise je Komplexitätsstufe (vgl. DIAS GmbH, 2020b). Aufgrund der hier zusammengefassten Unterschiede der Version vor März 2021 und der Version nach März 2021, kann keine im Rahmen dieser Arbeit getroffene Aussage ohne Weiteres auf den aktuellen BIK BITV-Test übertragen werden.

7.2 BITV-Audit

Beim BITV-Audit handelt es sich um ein Prüfverfahren für Web, Mobile, Dokumente und Desktop (vgl. T-Systems MMS, 2020a). Das Prüfverfahren wurde von der T-Systems MMS im Jahre 2009 zum ersten Mal verwendet. Im Jahr 2010 wurde der BITV-Audit erstmals

akkreditiert und wird seitdem kontinuierlich weiterentwickelt (A. Nebe [T-Systems MMS], persönliche Kommunikation, 12. April 2021). Derzeit wird beispielsweise daran gearbeitet das Prüfverfahren auch für Autorenwerkzeuge anwendbar zu machen (vgl. T-Systems MMS, o.D.c). Eine detaillierte Beschreibung des Prüfverfahrens inklusive der Prüfkriterien ist nicht veröffentlicht.

7.2.1 Zugrunde gelegter Standard

Dem Prüfverfahren zugrunde gelegt sind folgende Standards: BITV 2.0, EN 301 549, WCAG 2.1, EN ISO 9241-171[1] sowie ISO 14289-1[2] (vgl. T-Systems MMS, 2020b).

7.2.2 Varianten des Prüfverfahrens

Es gibt zwei Varianten des Audits, das vereinfachte Audit und das eingehende Audit. Mithilfe des vereinfachten Audits soll die IST-Konformität dargestellt werden. Hierfür werden Schwachstellen und deren Schweregrad mittels Problembeschreibungen sowie heuristischen Lösungsempfehlungen, die auf Normvorgaben basieren, in einem Kurzprotokoll festgehalten. Der eingehende Audit stellt im Auditprotokoll die SOLL-Konformität dar. Hierfür wird für jedes Problem zusätzlich eine Ursachenanalyse, eine Auswirkungsbeschreibung für jede Benutzergruppe und eine detaillierte Handlungsempfehlung angegeben. Dadurch soll die Problembehebung erleichtert werden. Beide Varianten führen vor der manuellen Prüfung eine automatisierte Prüfung durch. Optional kann auch noch eine Expertenevaluation der Usability durchgeführt werden (vgl. T-Systems MMS, 2020a). In den nachfolgenden Kapiteln wird jedoch ausschließlich Bezug auf das eingehende Audit genommen.

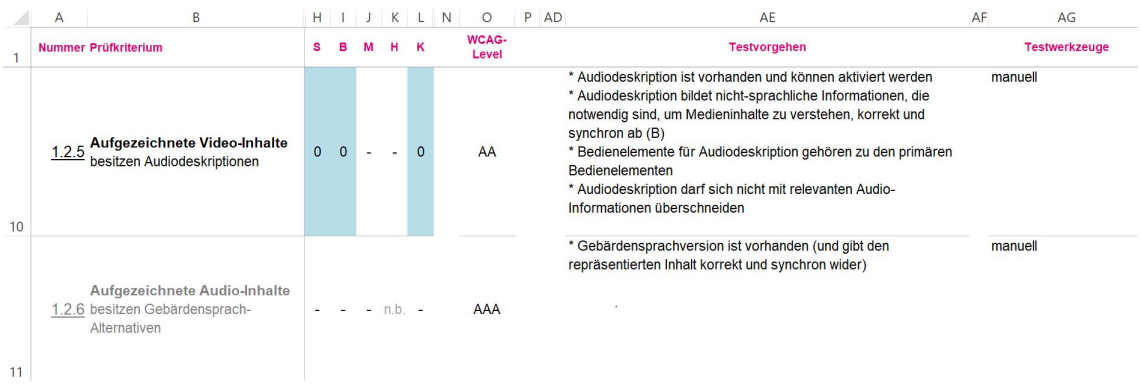
7.2.3 Prüfvorgehen

Abbildung 7.5 zeigt das Prüfvorgehen. Zu Beginn werden die Rahmenbedingungen in Zusammenarbeit mit dem Auftraggeber festgehalten. Auf dieser Grundlage wird zusammen die passende Variante des Prüfverfahrens ausgewählt und optional mit einem Usability-Test ergänzt. Schließlich muss der Prüfgegenstand bestimmt werden. Das heißt es wird festgelegt, welche Webseiten und Bereiche zu dem Webangebot dazu gehören. Die Komplexität sowie der Umfang des Prüfgegenstands sind ausschlaggebend für den Evaluationsaufwand

[1] Leitlinien für die Zugänglichkeit von Software (vgl. T-Systems MMS, 2020b)
[2] Verbesserung der Barrierefreiheit für das Dateiformat von elektronischen Dokumenten (vgl. T-Systems MMS, 2020b)

Abbildung 7.5: BITV-Audit: Schematische Darstellung des Prüfvorgehens

und den Preis. Der Prüfer erhält eine Programmeinführung durch den Auftraggeber. Bei einfachen Websites kann davon abgesehen werden. Auf Basis des Prüfgegenstands und den mit dem Auftraggeber getroffenen Vereinbarungen werden die relevanten Prüfkriterien ausgewählt. Es wird eine repräsentative Webseitenauswahl getroffen anhand derer die Prüfung durchgeführt wird. Die Prüfung besteht dann aus zwei Teilen, nämlich der Zugänglichkeitsprüfung und der Konformitätsprüfung. Bei ersterer handelt es sich um eine empirische Evaluation durch Experten, bei der die Barrierefreiheit des Prüfgegenstands aus der Perspektive von Benutzern mit Behinderung betrachtet wird. Hierfür werden fünf Benutzergruppen unterschieden: sehbeeinträchtigte, blinde, motorisch beeinträchtigte, hörgeschädigte sowie kognitiv beeinträchtigte Benutzer. Die durch die Expertenevaluation gefundenen Probleme werden anschließend den Konformitätskriterien zugeordnet. Schließlich werden auch noch nicht betrachtete Kriterien geprüft. Die Gewichtung der Probleme erfolgt nach Schwere der Auswirkungen auf die betroffenen Benutzergruppen. In einem Teamreview mit mindestens einem weiteren Prüfer werden alle gefundenen Probleme hinsichtlich ihrer Problemursache, Gewichtung und Handlungsempfehlung diskutiert. Danach erstellt der Prüfer den Prüfbericht. Dieser wird nochmals durch einen erfahrenen Prüfer formal abgenommen und an den Auftraggeber versendet (vgl. T-Systems MMS, 2020b).



	A	B	H	I	J	K	L	N	O	P	AD	AE	AF	AG
1	Nummer	Prüfkriterium	S	B	M	H	K		WCAG-Level			Testvorgehen		Testwerkzeuge
10	1.2.5	**Aufgezeichnete Video-Inhalte** besitzen Audiodeskriptionen	0	0	-	-	0		AA			* Audiodeskription ist vorhanden und können aktiviert werden * Audiodeskription bildet nicht-sprachliche Informationen, die notwendig sind, um Medieninhalte zu verstehen, korrekt und synchron ab (B) * Bedienelemente für Audiodeskription gehören zu den primären Bedienelementen * Audiodeskription darf sich nicht mit relevanten Audio-Informationen überschneiden		manuell
11	1.2.6	**Aufgezeichnete Audio-Inhalte** besitzen Gebärdensprach-Alternativen	-	-	-	n.b.	-		AAA			* Gebärdensprachversion ist vorhanden (und gibt den repräsentierten Inhalt korrekt und synchron wider)		manuell

Abbildung 7.6: BITV-Audit: Ausschnitt der Checkliste. Das Prüfkriterium 1.2.5 wird für drei Benutzergruppen bewertet. Das Prüfkriterium 1.2.6 wird nicht bewertet, da es dem WCAG Level AAA zugeordnet ist.

7.2.4 Toolunterstützung

Zur Unterstützung der Prüfer bei der Prüfungsdurchführung und der Prüfberichterstellung werden Dokumente genutzt, die mit Microsoft Office Produkten verarbeitet werden. Eine umfangreiche Checkliste (T-Systems MMS, o.D.c) enthält alle Prüfkriterien des BITV-Audits, die jeweilige Beschreibung des Prüfvorgehens sowie die anzuwendenden Testwerkzeuge. In einem ersten Schritt macht der Prüfer Angaben zum Prüfgegenstand, den zu betrachtenden Benutzergruppen und den anzuwendenden Standards. Dadurch werden automatisch die notwendigen Prüfkriterien ausgewählt und die nicht notwendigen Prüfkriterien in der Ansicht ausgegraut. Der Prüfer führt die empirische Evaluation durch und notiert alle auftretenden Probleme in einem separaten Arbeitsdokument. Anschließend kann er für jedes Prüfkriterium eine Bewertung für die betroffenen Benutzergruppen in der Checkliste festhalten, siehe Abbildung 7.6. Beim Teamreview können notwendige Anpassungen direkt in der Checkliste sowie dem Arbeitsdokument notiert werden. Zur anschließenden Erstellung des Prüfberichts wird ein Word Dokument (T-Systems MMS, o.D.e) genutzt. Dieses stellt mithilfe einer programmatischen Erweiterung wichtige Funktionen bereit, die über die üblichen Textverarbeitungsoptionen hinausgehen. Das Menü der verfügbaren Funktionen ist in Abbildung 7.7 zu sehen. Beispielsweise kann der Prüfer über die Schaltfläche „Dokument Eigenschaften“ alle notwendigen Informationen zum Prüfauftrag angeben. Die Prüfkriterien und die Bewertungen der einzelnen Benutzergruppen werden mithilfe der Funktion „Alle Kriterien einfügen“ automatisch aus der Checkliste extrahiert, in strukturierter Weise für die Weiterverarbeitung eingefügt und vorbereitet. Der Prüfer kann nun die noch fehlenden Informationen zu den im Teamreview besprochenen Problemen eintragen und Screenshots hinzufügen. Der Prüfer kann die Inhalte des Prüfberichts

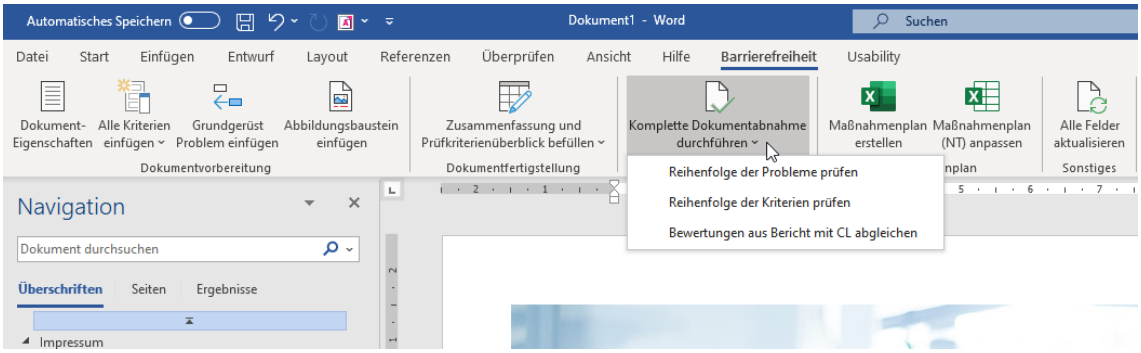


Abbildung 7.7: BITV-Audit: Menü der zusätzlichen Funktionen im Word-Dokument

mit den Informationen der Checkliste abgleichen lassen und schließlich aus dem Word Dokument ein PDF-Dokument für den Auftraggeber generieren lassen.

7.2.5 Prüferqualifikation

T-Systems MMS setzt auf einen kontinuierlichen internen Qualifizierungsprozess. Dabei wird unter anderem besonders Wert auf den sicheren Umgang mit Normen und Richtlinien gelegt. Aber auch die Prüfungskompetenz hinsichtlich der Planung und Durchführung von Barrierefreiheitsevaluationen steht im Fokus. Zudem werden die Experten im Erkennen und Bewerten der Auswirkungen von Barrierefreiheitsproblemen geschult. Wenn ein Prüfer gerade erst bei T-Systems MMS begonnen hat, unterstützt und begleitet ihn ein erfahrener Barrierefreiheitsprüfer in allen Projektphasen. Alle Barrierefreiheitsexperten erwerben folgende Zertifikate: ISTQB Certified Tester und UXQB Certified Professionals for Usability und User Experience. Mithilfe regelmäßiger Teamreviews von Prüfergebnissen und Besprechungen von Fachthemen in monatlichen Treffen soll die Qualifikation der Prüfer fortlaufend weiterentwickelt werden (vgl. T-Systems MMS, 2020b).

7.2.6 Prüfkriterien

Analog zu den Prüfschritten des BIK BITV-Tests gibt es die Prüfkriterien beim BITV-Audit. Insgesamt gibt es 85 Prüfkriterien, welche jedoch nicht öffentlich zugänglich sind. Die Auswahl der Prüfkriterien findet automatisiert vor Beginn des Tests statt und basiert auf dem gewünschten zugrunde gelegten Standard (z.B. WCAG 2.1 AA oder WCAG 2.1 AAA), den zu berücksichtigenden Benutzergruppen und dem jeweiligen Prüfgegenstand

Problemgewichtung	Definition
Leichte Zugänglichkeitseinschränkung	„[E]ine aufgabenrelevante Funktion [ist] nicht bedienbar oder eine aufgabenrelevante Information [ist] nicht wahrnehmbar [...].“
Zugänglichkeitshürde	„[E]ine aufgabenrelevante Information [ist] schwer verständlich oder schwer wahrnehmbar [...] bzw. eine aufgabenrelevante Funktion oder Information [kann] nur durch Umgehung des Problems mittels einer Hilfskonstruktion genutzt werden [...].“
Zugänglichkeitsblockade	„Informationen oder Funktionen [werden] erst durch bestimmte Benutzeraktionen zugänglich [...] bzw. der Benutzer [wird] gestört oder abgelenkt [...].“

Tabelle 7.2: BITV-Audit: Problemgewichtungen und ihre Definition (T-Systems MMS, 2020b)

(vgl. T-Systems MMS, 2020b). Um den gleichen Standard wie der BIK BITV-Test abdecken zu können, werden 50 der 85 Prüfkriterien benötigt (vgl. T-Systems MMS, o.D.c).

7.2.7 Bewertungskategorien

Beim BITV-Audit werden folgende drei Problemgewichtungen unterschieden: eine leichte Zugänglichkeitseinschränkung, eine Zugänglichkeitshürde sowie eine Zugänglichkeitsblockade. Tabelle 7.2 zeigt die Definitionen der drei Problemgewichtungen. Jedes gefundene Problem wird für die betroffenen Benutzergruppen jeweils einer der drei Problemgewichtungen zugeordnet (vgl. T-Systems MMS, 2020b).

Die Zugänglichkeitsaussage gibt an, wie gut das geprüfte Webangebot für die fünf Benutzergruppen zugänglich ist. Hierfür wird jeder Benutzergruppe eine Schulnote zugewiesen. Eine Aufschlüsselung der Schulnoten ist in Tabelle 7.3 zu sehen (vgl. T-Systems MMS, 2021a).

Die Konformitätsaussage wird wie folgt getroffen: Das Webangebot ist konform zur BITV 2.0, wenn es für alle fünf Benutzergruppen gut (2) oder sehr gut (1) zugänglich ist. Es ist teilweise konform zur BITV 2.0, wenn es für alle fünf Benutzergruppen mindestens eingeschränkt (3) oder besser zugänglich ist. Das Webangebot ist nicht konform zur BITV 2.0, wenn es für mindestens eine Benutzergruppe mangelhaft (4) oder ungenügend (5) zugänglich ist (vgl. T-Systems MMS, 2021a).

Zugänglichkeit	Leichte Zugänglichkeitseinschränkung	Zugänglichkeitshürde	Zugänglichkeitsblockade
1 sehr gut	0	0	0
2 gut	bis 3	0	0
3 eingeschränkt	ab 4	bis 3	0
4 mangelhaft	beliebig	ab 4	0
5 ungenügend	beliebig	beliebig	ab 1

Tabelle 7.3: BITV-Audit: Zugänglichkeitsaussage in Abhängigkeit von Problemgewichtungen (T-Systems MMS, 2021a)

7.2.8 Qualitätssicherung

Bei dem gemeinsamen Teamreview stellt der verantwortliche Prüfer die gefundenen Ergebnisse einem anderen Barrierefreiheitsprüfer vor. Gemeinsam werden dann die Aspekte Problemursache, Problemgewichtung und Handlungsempfehlungen diskutiert. Dadurch wird eine gleichbleibende Qualität der Ergebnisse gefördert und mögliche Bewertungsspielräume in ähnlicher Weise genutzt. Abschließend wird noch eine formale Abnahme durch einen erfahrenen Barrierefreiheitsexperten durchgeführt. Optional wird hier nochmals eine Korrekturschleife durchgeführt (vgl. T-Systems MMS, 2020b).

7.2.9 Prüfbericht

Der PDF-Prüfbericht gliedert sich in folgende Abschnitte: Management-Zusammenfassung, Problembeschreibungen und formale Angaben zur Prüfung. Die Management-Zusammenfassung enthält die Konformitätsbewertung des geprüften Webangebots. Sie definiert inwieweit das Webangebot konform zur BITV 2.0 ist. Zudem wird auch eine Zugänglichkeitsaussage für jede Benutzergruppe getroffen. Abbildung 7.8 zeigt jeweils einen Ausschnitt der Management-Zusammenfassung aus zwei Prüfberichten. Anschließend werden die verletzten Prüfkriterien aufgelistet, wobei die Problemgewichtung aufgeschlüsselt nach Benutzergruppen gezeigt wird (vgl. T-Systems MMS, 2020b).

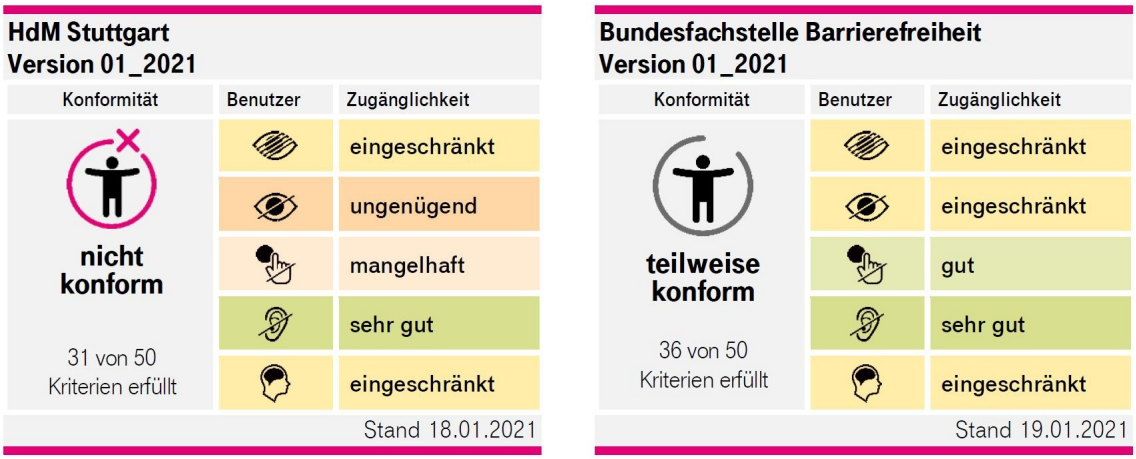


Abbildung 7.8: BITV-Audit: Ausschnitt der Management-Zusammenfassung aus zwei Prüfberichten

Die Problembeschreibungen umfassen die verletzten Prüfkriterien sowie alle dazugehörigen Probleme. Für jedes Problem werden folgende Aspekte dokumentiert: der Problemtitel, die Problemgewichtung für die betroffenen Benutzergruppen, eine Problembeschreibung mit Ursachenanalyse und Auswirkungsbeschreibung, einen Screenshot und eine Handlungsempfehlung. Der Prüfbericht endet mit den formalen Angaben zur Prüfung (vgl. T-Systems MMS, 2020b).

7.2.10 Veröffentlichung der Ergebnisse

Der PDF-Prüfbericht darf vom Auftraggeber veröffentlicht werden. Jedoch muss dann der Prüfbericht in vollem Umfang zugänglich gemacht werden. Es dürfen keine Teile weggelassen werden (A. Nebe [T-Systems MMS], persönliche Kommunikation, 01. März 2021).

7.2.11 Prüfzeichen

Es gibt drei verschiedene Prüfzeichen: „Nicht konform zur BITV 2.0“, „teilweise konform zur BITV 2.0“ und „konform zur BITV 2.0“. Das Prüfzeichen „Nicht konform zur BITV 2.0“ besagt, dass ein BITV-Audit durch die T-Systems MMS durchgeführt wurde, das Webangebot allerdings mindestens für eine Benutzergruppe mangelhaft oder ungenügend zugänglich ist. Das Prüfzeichen „teilweise konform zur BITV 2.0“ bedeutet, dass das Webangebot zumindest eingeschränkt zugänglich ist für alle Benutzergruppen. Das letzte Prüfzeichen „konform zur BITV 2.0“ besagt, dass die Zugänglichkeitsbewertung für jede

Komplexität	Erklärung	Preis
1	Einfache HTML-Seite - statisch, informationsorientiert	300 €
2	einfache dynamische Seite - Formulare, Videos, Karussell, Flyout-Menüs etc.	450 €
3	komplexe dynamische Seite - Webanwendungen, komplexe Formulare	600 €

Tabelle 7.4: BITV-Audit: Preis pro Webseite für verschiedene Komplexitätsstufen (T-Systems MMS, o.D.d)

Benutzergruppe mindestens gut oder sogar sehr gut ausgefallen ist. Das Prüfzeichen wird im Prüfbericht in der Management-Zusammenfassung abgebildet (vgl. T-Systems MMS, 2020b; T-Systems MMS, 2021a). Der Auftraggeber kann das Prüfzeichen auch auf der eigenen Website einbinden (A. Nebe [T-Systems MMS], persönliche Kommunikation, 01. März 2021).

7.2.12 Preise

Bei der Preisgestaltung wird unterschieden zwischen informationsorientierten und anwendungsorientierten Websites. Bei den informationsorientierten Websites wird der Preis mithilfe der in Tabelle 7.4 gezeigten Preisliste erstellt. Der Gesamtpreis entspricht der Summe aller Preise für die einzelnen zu prüfenden Webseiten. Bei anwendungsorientierten Websites wird workflow-basiert und nicht seitenbasiert geprüft. Hier wird eine individuelle Schätzung des Aufwands vorgenommen. Dieser Aufwand hängt von vielen verschiedenen Faktoren ab. Im Nachfolgenden sind einige dieser Einflussfaktoren genannt: Komplexität der Inhalte, eingesetzte (Web-)Technologien, Integration von Drittanbieterinhalten innerhalb eines zu prüfenden Workflows und unterschiedliche Ansichten je nach Rolle (A. Nebe [T-Systems MMS], persönliche Kommunikation, 12. April 2021).

8 Erstellter Kriterienkatalog

Abschnitt 8.1 beschreibt das Vorgehen zur Identifizierung und Gewichtung der Kriterien. In Abschnitt 8.2 werden grundlegende Informationen zu den Kriterien und Metriken gegeben, bevor dann in den nachfolgenden Abschnitten alle 22 Kriterien detailliert beschrieben und die verwendeten Metriken definiert werden. Die Abschnitte der Kriterien sind der Gewichtung nach geordnet, beginnend mit dem höchsten Gewicht.

8.1 Vorgehen

Vorab wurden 14 Kriterien durch die Literaturrecherche und ein Brainstorming mit den Betreuern dieser Arbeit definiert. Diese Vorauswahl wurde in einer Fokusgruppe mit Experten diskutiert, überarbeitet und mit weiteren Kriterien ergänzt. Die neunköpfige Expertenrunde bestehend aus Mitgliedern des BIK-Prüfverbunds, Mitarbeitern der T-Systems MMS sowie Barrierefreiheitsexperten traf sich in einer virtuellen Sitzung, die von der Autorin dieser Arbeit moderiert wurde. Da alle Experten den BIK BITV-Test kannten, war hier keine Einführung des Prüfverfahrens notwendig. Der BITV-Audit war bis dahin nur den drei Mitarbeitern der T-Systems MMS bekannt, daher wurde zu Beginn eine kurze Vorstellung des Prüfverfahrens durch eine Mitarbeiterin gegeben. Um eine Diskussionsgrundlage zu schaffen, wurden alle 14 Kriterien kurz vorgestellt. In einer 60-minütigen Diskussion im Plenum wurde versucht ein gemeinsames Verständnis der bisherigen Kriterien zu schaffen, aber auch weitere notwendige Kriterien zu identifizieren. Schließlich wurden insgesamt 22 Kriterien festgelegt. Am Ende der Expertenrunde wurde das Vorgehen zur Gewichtung der einzelnen Kriterien von der Moderatorin erklärt.

Im Anschluss an die virtuelle Expertenrunde wurde jedem Teilnehmer eine Excel-Datei mit den 22 Kriterien zugesendet. Diese Excel-Datei ist im elektronischen Anhang zu finden. Hiermit konnten alle Experten ihre persönliche Gewichtung vornehmen. Sie durften für jedes Kriterium beliebig viele Punkte vergeben. Die Vorgabe hierfür war jedoch, dass ein Kriterium mehr Punkte erhielt, desto wichtiger es für sie war. Falls ein Kriterium als

vollkommen irrelevant angesehen wurde, war auch die Vergabe von 0 Punkten möglich. Insgesamt haben acht der neun Experten eine Gewichtung vorgenommen. Die Gewichtungen wurden jeweils normalisiert und dann gemittelt. Es konnten keine systematischen Unterschiede der Gewichtung zwischen den Teilnehmern mit vorheriger Kenntnis des zweiten Prüfverfahrens und denen ohne festgestellt werden. Die einzelnen Expertengewichtungen sowie die daraus gemittelte Gewichtung sind im elektronischen Anhang zu finden.

8.2 Kriterien und Metriken

$$\sum_{kriterium=1}^{22} Gewicht_{kriterium} = 1 \quad mit\ 0 \leq Gewicht \leq 1$$

Die Summe der Gewichte aller 22 Kriterien ist immer 1. Der Gesamtscore eines Prüfverfahrens entspricht der Summe des Produkts aus Kriteriumscores und ihren jeweiligen Gewichten. Dabei können der Gesamtscore, Kriteriumscore und die Gewichte nur Werte von 0 bis 1 annehmen.

$$Gesamtscore = \sum_{kriterium=1}^{22} \left(Kriteriumscore_{kriterium} \cdot Gewicht_{kriterium} \right)$$

$$mit\ 0 \leq Gesamtscore, Kriteriumscore, Gewicht \leq 1$$

Ein Kriteriumscore entspricht dem Mittelwert der jeweiligen Metrikscores. Ein Kriterium kann durch bis zu sechs Metriken definiert werden. Sollte ein Kriterium aus mehreren Metriken berechnet werden, werden alle Metriken gleich gewichtet. Auch der Metrikscore kann nur einen Wert von 0 bis 1 annehmen. Bei allen Metriken wird eine genaue Berechnung angestrebt, da die Gewichtung der Experten ein Multiplikator der Differenz sein soll. Das heißt, sollte beispielsweise Prüfverfahren A doppelt so viele Anforderungen eines Standards X abdecken wie Prüfverfahren B, soll sich dies auch in den Metrikscores der Prüfverfahren widerspiegeln. Eine einfache Mehrheitsentscheidung wurde abgelehnt, wie Prüfverfahren A erhält den Wert 1 und Prüfverfahren B den Wert 0, weil A mehr Anforderungen des Standards X abdeckt als B. Eine Übersicht der 22 Kriterien und 41 Metriken ist im elektronischen Anhang zu finden.

$$Kriteriumscore_{kriterium} = \frac{1}{M_{kriterium}} \sum_{metrik=1}^{M_{kriterium}} Metrikscore_{kriterium,metrik}$$

$$mit\, 0 \leq Kriteriumscore, Metrikscore \leq 1\, und\, M \in \{1; 2; 3; 6\}$$

Notwendige Normalisierungen von Werten werden anhand folgender Gleichungen durchgeführt:

$$Norm = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad Norm_{rev} = 1 - \left(\frac{x - x_{min}}{x_{max} - x_{min}}\right)$$

8.3 Abdeckung von WCAG 2.1 AA

Aufgrund der Expertenrunde wurde für dieses Kriterium ein Gewicht von 7,72% errechnet. Das Kriterium wird anhand der Metrik „Anzahl der WCAG 2.1 Erfolgskriterien mit Konformitätsstufe A und AA, die vom Prüfverfahren abgedeckt werden“ bewertet.

$$Abdeckung_{\,WCAG\,2.1\,AA} = \frac{1}{50} \sum_{kriterium=1}^{50} x_{kriterium} \qquad mit\, x \in \{0; 1\}$$

Es wird jedem der 50 WCAG 2.1 Erfolgskriterien mit Konformitätsstufe A oder AA der Wert 0 („nicht abgedeckt“) oder 1 („abgedeckt“) zugewiesen und dann der Mittelwert über alle Erfolgskriterien gebildet (vgl. W3C, 2018). Es gilt: Je mehr Erfolgskriterien vom Prüfverfahren abgedeckt werden, desto besser ist die Abdeckung von WCAG 2.1 AA und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibungen der zu vergleichenden Prüfverfahren durchgeführt werden.

8.4 Vollständigkeit

Aufgrund der Expertenrunde wurde für dieses Kriterium ein Gewicht von 7,39% errechnet. Das Kriterium wird anhand der beiden Metriken „Vollständigkeit (Bewertung nach WCAG 2.1 AA)“ und „Vollständigkeit (Bewertung aus Nutzersicht)“ bewertet. Die Notwendigkeit dieser Differenzierung der Bewertung hatte sich in einem Workshop mit Experten herauskristallisiert. Es zeigte sich, dass die Einschätzung, ob es sich bei einem im Prüfbericht

dokumentierten Problem auch tatsächlich um ein Problem handelt, für diese beiden Bewertungsblickwinkel unterschiedlich ausfallen kann.

$$Vollständigkeit_{WCAG} = \frac{1}{W} \sum_{website=1}^{W} \frac{KP_{website}}{KP_{website} + FN_{website}} \quad mit\, W \in \mathbb{N}$$

$$Vollständigkeit_{Nutzer} = \frac{1}{W} \sum_{website=1}^{W} \frac{KP_{website}}{KP_{website} + FN_{website}} \quad mit\, W \in \mathbb{N}$$

Die Vollständigkeit entspricht bei beiden Metriken dem Verhältnis der korrekt positiven Ergebnisse (KP) und der falsch negativen Ergebnisse (FN). Bei einem korrekt positiven Ergebnis handelt es sich um ein Problem, das vom Prüfverfahren korrekterweise als Problem identifiziert wird, da es nach WCAG 2.1 AA bzw. aus Nutzersicht ein Problem darstellt. Bei einem falsch negativen Ergebnis handelt es sich um ein Problem, das vom Prüfverfahren fälschlicherweise nicht als Problem identifiziert wird, obwohl es nach WCAG 2.1 AA bzw. aus Nutzersicht ein Problem darstellt. Zur Berechnung der Vollständigkeit (vgl. Abduganiev, 2017) werden die KP und FP für jede Website aus beiden Bewertungsblickwinkeln ermittelt und wie in den oben definierten Gleichungen ins Verhältnis gesetzt. Es gilt: Je weniger falsch negative Ergebnisse, desto besser ist die Vollständigkeit und desto höher ist der normalisierte Wert der Metrik. Die Auswertung kann anhand der Prüfberichtdaten der zu vergleichenden Prüfverfahren durchgeführt werden.

8.5 Qualität des Prüfberichts

Aufgrund der Expertenrunde wurde für dieses Kriterium ein Gewicht von 6,67% errechnet. Das Kriterium wird anhand der sechs Metriken „Verfügbarkeit von Screenshots“, „Verfügbarkeit von Markierungen in Screenshots“, „Verfügbarkeit von dokumentierten Lösungen zu den Barrieren“, „Anpassbarkeit der Ausführlichkeit“, „Verständlichkeit für den Leser hinsichtlich Verstehen der Barrieren“ und „Barrierefreiheit des Prüfberichts“ bewertet.

8.5.1 Verfügbarkeit von Screenshots

$$Verfügbarkeit_{Screenshots}$$

$$= \frac{1}{W} \sum_{website=1}^{W} \frac{Anzahl_Probleme_mit_Screenshot_{website}}{Anzahl_Probleme_{website}} \quad mit\, W \in \mathbb{N}$$

Für die Metrik „Verfügbarkeit von Screenshots“ wird für jede Website die Anzahl der dokumentierten Probleme mit Screenshots geteilt durch die Gesamtzahl der dokumentierten Probleme. Anschließend wird der Mittelwert über alle Websites gebildet. Es gilt: Je mehr Screenshots enthalten sind, desto besser ist die Qualität des Prüfberichts und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Prüfberichtdaten der zu vergleichenden Prüfverfahren durchgeführt werden.

8.5.2 Verfügbarkeit von Markierungen in Screenshots

$$Verfügbarkeit_{Markierungen}$$

$$= \frac{1}{W} \sum_{website=1}^{W} \frac{Anzahl_Probleme_mit_Markierung_{website}}{Anzahl_Probleme_{website}} \quad mit\, W \in \mathbb{N}$$

Für die Metrik „Verfügbarkeit von Markierungen in Screenshots“ wird für jede Website die Anzahl der dokumentierten Probleme mit Markierungen in Screenshots geteilt durch die Gesamtzahl der dokumentierten Probleme. Anschließend wird der Mittelwert über alle Websites gebildet. Es gilt: Je mehr Screenshots mit Markierungen enthalten sind, desto besser ist die Qualität des Prüfberichts und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Prüfberichtdaten der zu vergleichenden Prüfverfahren durchgeführt werden.

8.5.3 Verfügbarkeit von dokumentierten Lösungen zu den Barrieren

$$Verfügbarkeit_{Lösungen}$$

$$= \frac{1}{W} \sum_{website=1}^{W} \frac{Anzahl_Probleme_mit_Empfehlung_{website}}{Anzahl_Probleme_{website}} \quad mit\, W \in \mathbb{N}$$

Für die Metrik „Verfügbarkeit von dokumentierten Lösungen zu den Barrieren“ wird für jede Website die Anzahl der dokumentierten Probleme mit Empfehlung zur Problemlösung geteilt durch die Gesamtzahl der dokumentierten Probleme. Anschließend wird der Mittelwert über alle Websites gebildet. Es gilt: Je mehr Empfehlungen enthalten sind, desto besser ist die Qualität des Prüfberichts und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Prüfberichtdaten der zu vergleichenden Prüfverfahren durchgeführt werden.

8.5.4 Anpassbarkeit der Ausführlichkeit

$$Anpassbarkeit = \frac{1}{2} \sum_{aspekt=1}^{2} x_{aspekt} \quad mit\, x \in \{0; 1\}$$

Für die Metrik „Anpassbarkeit der Ausführlichkeit“ werden folgende zwei Aspekte mit dem Wert 0 („falsch“) oder 1 („korrekt“) bewertet:

- Es gibt eine Ergebniszusammenfassung.
- Es gibt eine Alternativversion mit geringerem oder höherem Umfang.

Anschließend wird der Mittelwert über beide Aspekte gebildet. Es gilt: Je mehr Aspekte mit 1 („korrekt“) bewertet werden, desto besser ist die Anpassbarkeit der Ausführlichkeit, desto besser ist die Qualität des Prüfberichts und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibungen der zu vergleichenden Prüfverfahren durchgeführt werden.

8.5.5 Verständlichkeit für den Leser hinsichtlich Verstehen der Barrieren

$$Verständlichkeit$$

$$= Norm\left(\frac{1}{I} \sum_{interview=1}^{I} x_{interview}\right) \quad mit\, I \in \mathbb{N}\, und\, x \in \{1; 2; ...; 6; 7\}$$

Für die Metrik „Verständlichkeit für den Leser hinsichtlich Verstehen der Barrieren“ wird folgende Interviewfrage betrachtet: „Inwieweit trägt der Prüfbericht für die Leser zum Verständnis der gefundenen Barrieren bei?“. Diese Frage wird anhand einer Likert-Skala von 1 („überhaupt nicht“) bis 7 („vollständig“) bewertet und ein Mittelwert über alle Interviews gebildet. Anschließend wird dieser Mittelwert normalisiert. Zur Normalisierung der Werte wird als minimaler Wert 1 und als maximaler Wert 7 genommen. Es gilt: Je höher die Wertung der Interviewpartner ist, desto besser ist die Verständlichkeit, desto besser ist die Qualität des Prüfberichts und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand von Interviewdaten durchgeführt werden.

8.5.6 Barrierefreiheit des Prüfberichts

$$Barrierefreiheit_{Prüfbericht}$$

$$= \frac{1}{W} \sum_{website=1}^{W} \left(\frac{1}{11} \sum_{prüfpunkt=1}^{11} Norm_{rev}(Anzahl_Fehler_{website,prüfpunkt}) \right)$$

$$mit\, W \in \mathbb{N}$$

Für die Metrik „Barrierefreiheit des Prüfberichts“ werden die Prüfberichte der geprüften Websites mit dem PDF Accessibility Checker[1] Version 3 (PAC 3) getestet. Das Tool wurde auf Basis einer umfassenden Recherche zu automatisierten PDF-Prüftools ausgewählt. Die Auswertung des PAC 3 umfasst insgesamt elf Prüfpunkte. Für jeden Prüfpunkt werden die in den Prüfberichten gefundenen Fehler (mit dem Label „nicht erfüllt“) normalisiert. Zur Normalisierung der Werte wird für jeden Prüfpunkt als minimaler Wert 0 und als maximaler Wert die von allen zu vergleichenden Prüfverfahren höchste Anzahl der gefundenen Fehler dieses Prüfpunkts genommen. Dann wird der Mittelwert über alle Prüfpunkte gebildet und anschließend wird der Mittelwert über alle Websites berechnet. Es gilt: Je weniger Probleme gefunden werden, desto besser ist der Grad der Barrierefreiheit des Prüfberichts, desto besser ist die Qualität des Prüfberichts und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Prüfberichtdaten der zu vergleichenden Prüfverfahren durchgeführt werden.

8.6 Abdeckung von zusätzlichen Kriterien aus EN 301 549 Tab. A.1

Aufgrund der Expertenrunde wurde für dieses Kriterium ein Gewicht von 6,64% errechnet. Das Kriterium wird anhand der Metrik „Anzahl der zusätzlichen Kriterien der EN 301 549 Tab. A.1, die vom Prüfverfahren abgedeckt werden“ bewertet. Als Grundlage werden alle in Tabelle A.1 aus EN 401 549 V3.2.1 (Europäisches Institut für Telekommunikationsnormen, 2020) genannten Kriterien betrachtet, die nicht durch die WCAG 2.1 AA abgedeckt werden.

[1] Kostenloses Tool von Zugang für alle unter `https://www.access-for-all.ch/ch/pdf-werkstatt/pdf-accessibility-checker-pac.html` verfügbar.

$$Abdeckung_{EN\ 301\ 549} = \frac{1}{42} \sum_{kriterium=1}^{42} x_{kriterium} \quad mit\ x \in \{0;1\}$$

Es wird jedem der 42 Kriterien der Wert 0 („nicht abgedeckt“) oder 1 („abgedeckt“) zugewiesen und dann der Mittelwert über alle Kriterien gebildet. Es gilt: Je mehr zusätzliche Kriterien vom Prüfverfahren abgedeckt werden, desto besser ist die Abdeckung der EN 301 549 und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibungen der zu vergleichenden Prüfverfahren durchgeführt werden.

8.7 Korrektheit

Aufgrund der Expertenrunde wurde für dieses Kriterium ein Gewicht von 6,32% errechnet. Das Kriterium wird anhand der beiden Metriken „Korrektheit (Bewertung nach WCAG 2.1 AA)“ und „Korrektheit (Bewertung aus Nutzersicht)“ bewertet. Die Notwendigkeit dieser Differenzierung der Bewertung hatte sich in einem Workshop mit Experten herauskristallisiert. Es zeigte sich, dass die Einschätzung, ob es sich bei einem im Prüfbericht dokumentierten Problem auch tatsächlich um ein Problem handelt, für diese beiden Bewertungsblickwinkel unterschiedlich ausfallen kann.

$$Korrektheit_{WCAG} = \frac{1}{W} \sum_{website=1}^{W} \frac{KP_{website}}{KP_{website} + FP_{website}} \quad mit\ W \in \mathbb{N}$$

$$Korrektheit_{Nutzer} = \frac{1}{W} \sum_{website=1}^{W} \frac{KP_{website}}{KP_{website} + FP_{website}} \quad mit\ W \in \mathbb{N}$$

Die Korrektheit entspricht bei beiden Metriken dem Verhältnis der korrekt positiven Ergebnisse (KP) und der falsch positiven Ergebnisse (FP). Bei einem korrekt positiven Ergebnis handelt es sich um ein Problem, das vom Prüfverfahren korrekterweise als Problem identifiziert wird, da es nach WCAG 2.1 AA bzw. aus Nutzersicht ein Problem darstellt. Bei einem falsch positiven Ergebnis handelt es sich um ein Problem, das vom Prüfverfahren fälschlicherweise als Problem identifiziert wird, obwohl es nach WCAG 2.1 AA bzw. aus Nutzersicht kein Problem darstellt. Zur Berechnung der Korrektheit (vgl. Abduganiev, 2017) werden die KP und FP für jede Website aus beiden Bewertungsblickwinkeln ermittelt und wie in den oben definierten Gleichungen ins Verhältnis gesetzt. Es gilt: Je weniger

falsch positive Ergebnisse auftreten, desto besser ist die Korrektheit und desto höher ist der normalisierte Wert der Metrik. Die Auswertung kann anhand der Prüfberichtdaten der zu vergleichenden Prüfverfahren durchgeführt werden.

8.8 Qualität der Stichprobe

Aufgrund der Expertenrunde wurde für dieses Kriterium ein Gewicht von 5,96% errechnet. Das Kriterium wird anhand der beiden Metriken „Umfang der Stichprobe“ und „Varianz der Stichprobe“ bewertet. Bei der Stichprobe handelt es sich um die Auswahl der Webseiten einer Website, die geprüft werden soll.

8.8.1 Umfang der Stichprobe

$$Umfang_{Stichprobe}$$

$$= \frac{1}{3}\left(x + \sum_{variante=1}^{2} Norm(Anzahl_Webseiten_{variante})\right) \quad mit\ x \in \{0;1\}$$

Für die Metrik „Umfang der Stichprobe“ wird dem Aspekt „Die Anzahl der Webseiten ist variabel je nach Komplexität.“ der Wert 0 („falsch“) oder 1 („korrekt“) zugeordnet. Zudem wird sowohl die minimale Anzahl von Webseiten als auch die maximale Anzahl von Webseiten normalisiert. Der Mittelwert dieser drei Werte ergibt den Wert der Metrik. Diese drei Einflussgrößen wurden aufgrund einer Analyse der Verfahrensbeschreibungen vom BIK BITV-Test (DIAS GmbH, 2019) und dem BITV-Audit (T-Systems MMS, 2020b) sowie dem Anhang 1 Absatz 3 des Durchführungsbeschluss EU 2018/1524 (Europäische Kommission, 2018) ausgewählt. Es gilt: Je mehr Webseiten ausgewählt werden und die Komplexität berücksichtigt wird, desto größer ist der Umfang der Stichprobe, desto besser ist die Qualität der Stichprobe und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibungen der zu vergleichenden Prüfverfahren durchgeführt werden.

8.8.2 Varianz der Stichprobe

$$Varianz_{Stichprobe} = \frac{1}{10} \sum_{aspekt=1}^{10} x_{aspekt} \quad mit\, x \in \{0; 1\}$$

Für die Metrik „Varianz der Stichprobe“ werden folgende zehn Aspekte mit dem Wert 0 („falsch“) oder 1 („korrekt“) bewertet:

- Auswahl ist unabhängig vom Auftraggeber.
- Auswahl ist nicht vorhersehbar.
- Auswahl erfasst sämtliche Barrieren.
- Auswahl umfasst unterschiedliche Webseitentypen.
- Auswahl umfasst unterschiedliche Webseitenzustände.
- Auswahl deckt ganze Prozesse ab.
- Auswahl berücksichtigt Stellenwert einer Webseite für Benutzbarkeit des Webangebots.
- Auswahl umfasst unterschiedliche Funktionen.
- Auswahl nutzt das Zufallsprinzip.
- Auswahl umfasst Nicht-Web-Dokumente.

Anschließend wird der Mittelwert über alle Aspekte gebildet. Die oben genannten Aspekte wurden aufgrund einer Analyse der Verfahrensbeschreibungen vom BIK BITV-Test (DIAS GmbH, 2019) und dem BITV-Audit (T-Systems MMS, 2020b) sowie dem Anhang 1 Absatz 3 des Durchführungsbeschluss EU 2018/1524 (Europäische Kommission, 2018) ausgewählt. Es gilt: Je mehr Aspekte mit 1 („korrekt“) bewertet werden, desto besser ist die Varianz der Stichprobe, desto besser ist die Qualität der Stichprobe und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibungen der zu vergleichenden Prüfverfahren durchgeführt werden.

8.9 Qualitätssicherung

Aufgrund der Expertenrunde wurde für dieses Kriterium ein Gewicht von 5,58% errechnet. Das Kriterium wird anhand der drei Metriken „Qualitätssicherung innerhalb eines Prüfauftrags“, „Sicherstellung der Einheitlichkeit der Prüfung“ und „Sicherstellung der Qualifikation eines Prüfers“ bewertet.

8.9.1 Qualitätssicherung innerhalb eines Prüfauftrags

$$Qualitätssicherung_{Prüfauftrag} = \frac{1}{5} \sum_{aspekt=1}^{5} x_{aspekt} \quad mit\ x \in \{0;1\}$$

Für die Metrik „Qualitätssicherung innerhalb eines Prüfauftrags“ werden folgende fünf Aspekte mit dem Wert 0 („falsch“) oder 1 („korrekt“) bewertet:

- Die Webseitenauswahl (Stichprobe) wird von mehreren Experten erstellt bzw. überprüft.
- Mehrere Prüfer führen die Prüfung unabhängig voneinander durch.
- Es gibt eine Teambesprechung der gefundenen Probleme.
- Es gibt eine Qualitätssicherung der Prüfergebnisse durch eine qualifizierte Stelle.
- Der gleiche Prüfer prüft kein identisches Webangebot mehrfach.

Anschließend wird der Mittelwert über alle Aspekte gebildet. Die oben genannten Aspekte wurden aufgrund einer Analyse der Verfahrensbeschreibungen vom BIK BITV-Test (DIAS GmbH, 2019) und dem BITV-Audit (T-Systems MMS, 2020b) ausgewählt. Es gilt: Je mehr Aspekte mit 1 („korrekt“) bewertet werden, desto besser ist die Qualitätssicherung innerhalb eines Prüfauftrags, desto besser ist die Qualitätssicherung und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibungen der zu vergleichenden Prüfverfahren durchgeführt werden.

8.9.2 Sicherstellung der Einheitlichkeit der Prüfung

$$Sicherstellung_{Einheitlichkeit} = \frac{1}{4} \sum_{aspekt=1}^{4} x_{aspekt} \quad mit\ x \in \{0;1\}$$

Für die Metrik „Sicherstellung der Einheitlichkeit der Prüfung“ werden folgende vier Aspekte mit dem Wert 0 („falsch“) oder 1 („korrekt“) bewertet:

- Die Prüfer haben Zugriff auf veröffentlichte Prüfergebnisse.
- Die Prüfer haben Zugriff auf alle Prüfergebnisse.
- Eine Teambesprechung der gefundenen Probleme ist ein fester Prozessschritt der Prüfung.
- Es gibt eine statistische Auswertung der Einheitlichkeit.

Anschließend wird der Mittelwert über alle Aspekte gebildet. Die oben genannten Aspekte wurden aufgrund einer Analyse der Verfahrensbeschreibungen vom BIK BITV-Test (DIAS GmbH, 2019) und dem BITV-Audit (T-Systems MMS, 2020b) ausgewählt. Es gilt: Je mehr Aspekte mit 1 („korrekt“) bewertet werden, desto besser ist die Sicherstellung der Einheitlichkeit der Prüfung, desto besser ist die Qualitätssicherung und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibungen der zu vergleichenden Prüfverfahren durchgeführt werden.

8.9.3 Sicherstellung der Qualifikation eines Prüfers

$$Sicherstellung_{Prüferqualifikation} = \frac{1}{7} \sum_{aspekt=1}^{7} x_{aspekt} \quad mit\ x \in \{0; 1\}$$

Für die Metrik „Sicherstellung der Qualifikation eines Prüfers“ werden folgende sieben Aspekte mit dem Wert 0 („falsch“) oder 1 („korrekt“) bewertet:

- Es gibt einen Qualifizierungsprozess, um qualifizierter Prüfer zu werden.
- Gerade erst qualifizierte Prüfer werden zu Beginn von einem erfahrenen Prüfer begleitet.
- Es gibt eine periodische Prüfung der Qualifikation eines Prüfers.
- Es gibt regelmäßige Fortbildungsmaßnahmen bzw. Workshops.
- Es muss eine festgelegte Anzahl von Tests pro Jahr durchgeführt werden.
- Es gibt eine statistische Auswertung der Qualifikation eines Prüfers.
- Es gibt einen regelmäßigen Austausch mit Menschen mit Behinderung.

Anschließend wird der Mittelwert über alle Aspekte gebildet. Die oben genannten Aspekte wurden aufgrund einer Analyse der Verfahrensbeschreibungen vom BIK BITV-Test (DIAS GmbH, 2019) und dem BITV-Audit (T-Systems MMS, 2020b) ausgewählt. Es gilt: Je mehr Aspekte mit 1 („korrekt“) bewertet werden, desto besser ist die Sicherstellung der Qualifikation eines Prüfers, desto besser ist die Qualitätssicherung und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibungen der zu vergleichenden Prüfverfahren durchgeführt werden.

8.10 Optionale Inputformate

Aufgrund der Expertenrunde wurde für dieses Kriterium ein Gewicht von 5,57% errechnet. Das Kriterium wird anhand der Metrik „Anzahl der optionalen Inputformate“ bewertet.

$$Anzahl_{Inputformate} = \frac{1}{6} \sum_{dateiformat=1}^{6} x_{dateiformat} \quad mit\ x \in \{0;1\}$$

Es wird jedem der sechs Dateiformate der Wert 0 („nicht abgedeckt“) oder 1 („abgedeckt“) zugewiesen und dann der Mittelwert über alle Dateiformate gebildet. Die folgenden Dateiformate werden berücksichtigt: PDF-Dokument, Textverarbeitungsdokument, Tabellenkalkulationsdokument, Präsentationsdatei, Videodatei und Audiodatei. Sie wurden durch eine Recherche zu bekannten Dateiformaten, die über das Web geteilt werden, identifiziert und in die genannten Typen gruppiert. Website wird hier nicht aufgeführt, da bei Prüfmethoden zur Bewertung der Barrierefreiheit von Websites davon ausgegangen werden kann, dass dieses Format unterstützt wird. Es gilt: Je mehr optionale Dateiformate vom Prüfverfahren abgedeckt werden, desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibungen der zu vergleichenden Prüfverfahren durchgeführt werden.

8.11 Abdeckung von zusätzlichen Kriterien aus BITV 2.0

Aufgrund der Expertenrunde wurde für dieses Kriterium ein Gewicht von 5,53% errechnet. Das Kriterium wird anhand der Metrik „Anzahl der zusätzlichen Kriterien der BITV 2.0, die vom Prüfverfahren abgedeckt werden“ bewertet.

$$Abdeckung_{BITV\ 2.0} = \frac{1}{4} \sum_{kriterium=1}^{4} x_{kriterium} \quad mit\ x \in \{0;1\}$$

Es werden in Paragraph 4 und 7 der BITV 2.0 noch folgende zusätzliche Kriterien eingeführt: EzB, Feedback-Mechanismus, Erläuterungen in Leichter Sprache und Deutscher Gebärdensprache (vgl. Bundesministerium für Arbeit und Soziales, 2019b). Es wird jedem der vier Kriterien der Wert 0 („nicht abgedeckt“) oder 1 („abgedeckt“) zugewiesen und dann der Mittelwert über alle Kriterien gebildet. Es gilt: Je mehr zusätzliche Kriterien vom Prüfverfahren abgedeckt werden, desto besser ist die Abdeckung der BITV 2.0 und

desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibungen der zu vergleichenden Prüfverfahren durchgeführt werden.

8.12 Aufwand

Aufgrund der Expertenrunde wurde für dieses Kriterium ein Gewicht von 4,77% errechnet. Das Kriterium wird anhand der drei Metriken „Anzahl der Prüfschritte/-kriterien“, „Anzahl der involvierten Experten“ und „Zeitaufwand für Prüfprozess“ bewertet.

8.12.1 Anzahl der Prüfschritte/-kriterien

$$Aufwand_{Prüfschritte/-kriterien} = Norm_{rev}(Anzahl_Prüfschritte)$$

Für die Metrik „Anzahl der Prüfschritte/-kriterien“ musste ein gemeinsamer Standard ausgewählt werden, um die Prüfschritte/-kriterien, falls vom Prüfverfahren möglich, anzupassen. Es wurde sich für WCAG 2.1 AA als Basis entschieden, da dies der größte gemeinsame Nenner der im Rahmen dieser Arbeit zu vergleichenden Prüfverfahren ist. Zudem hat die Abdeckung dieser Anforderungen das höchste Gewicht bei der Expertengewichtung erhalten. Der Wert der Metrik entspricht der normalisierten Anzahl der dafür notwendigen Prüfschritte/-kriterien. Zur Normalisierung der Werte wird als minimaler Wert 1 und als maximaler Wert die von allen zu vergleichenden Prüfverfahren höchste Anzahl der Prüfschritte/-kriterien genommen. Es gilt: Je weniger Prüfschritte/-kriterien benötigt werden, desto geringer ist der Aufwand und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibungen der zu vergleichenden Prüfverfahren durchgeführt werden.

8.12.2 Anzahl der involvierten Experten

$$Aufwand_{Experten} = \frac{1}{2} \sum_{phase=1}^{2} Norm_{rev}(Anzahl_Experten_{phase})$$

Für die Metrik „Anzahl der involvierten Experten“ wird für die beiden Phasen „Prüfung“ und „Qualitätssicherung“ die Anzahl der benötigten Experten normalisiert. Zur Normalisierung der Werte wird für jede Phase als minimaler Wert 1 und als maximaler Wert die

von allen zu vergleichenden Prüfverfahren höchste Anzahl der Experten dieser Phase genommen. Anschließend wird der Mittelwert über beide Phasen gebildet. Es gilt: Je weniger Experten benötigt werden, desto geringer ist der Aufwand und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibungen der zu vergleichenden Prüfverfahren durchgeführt werden.

8.12.3 Zeitaufwand für Prüfprozess

$$Aufwand_{Zeit} = \frac{1}{W} \sum_{website=1}^{W} Norm_{rev}(Benötigte_Zeit_{website}) \quad mit\, W \in \mathbb{N}$$

Für die Metrik „Zeitaufwand für Prüfprozess“ wird für jede Website die für die Prüfung (inklusive der Erstellung des Prüfberichts und Qualitätssicherung) benötigte Zeit normalisiert. Diese muss bei der Durchführung der Prüfungen genau dokumentiert werden. Zur Normalisierung der Werte wird für jede Website als minimaler Wert 0 und als maximaler Wert die von allen zu vergleichenden Prüfverfahren höchste benötigte Zeit dieser Website genommen. Anschließend wird der Mittelwert über alle Websites gebildet. Es gilt: Je weniger Zeit benötigt wird, desto geringer ist der Aufwand und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der dokumentierten Zeitmessung bei der Erstellung der Prüfberichtdaten durchgeführt werden.

8.13 Abdeckung von zusätzlichen Kriterien aus WCAG 2.1 AAA

Aufgrund der Expertenrunde wurde für dieses Kriterium ein Gewicht von 4,70% errechnet. Das Kriterium wird anhand der Metrik „Anzahl der WCAG 2.1 Erfolgskriterien der Konformitätsstufe AAA, die vom Prüfverfahren abgedeckt werden“ bewertet.

$$Abdeckung_{\,WCAG\,2.1\,AAA} = \frac{1}{28} \sum_{kriterium=1}^{28} x_{kriterium} \quad mit\, x \in \{0;1\}$$

Es wird jedem der 28 WCAG 2.1 Erfolgskriterien mit Konformitätsstufe AAA der Wert 0 („nicht abgedeckt“) oder 1 („abgedeckt“) zugewiesen und dann der Mittelwert über alle Erfolgskriterien gebildet (vgl. W3C, 2018). Es gilt: Je mehr Erfolgskriterien vom Prüfverfahren abgedeckt werden, desto besser ist die Abdeckung von WCAG 2.1 AAA und

desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibungen der zu vergleichenden Prüfverfahren durchgeführt werden.

8.14 Bandbreite an geprüften Benutzungskonstellationen

Aufgrund der Expertenrunde wurde für dieses Kriterium ein Gewicht von 4,19% errechnet. Das Kriterium wird anhand der drei Metriken „Verwendete Plattformen“, „Verwendete Browser“ und „Verwendete Hilfsmittel“ bewertet.

8.14.1 Verwendete Plattformen

$$Abdeckung_{\,Plattformen} = \frac{1}{5} \sum_{plattform=1}^{5} x_{plattform} \quad mit\ x \in \{0; 0,5; 1\}$$

Für die Metrik „Verwendete Plattformen“ wird jedem der fünf Plattformen der Wert 0 („in Prüfung nicht enthalten“), 0,5 („in Prüfung optional auf Wunsch des Auftraggebers enthalten“) oder 1 („in Prüfung enthalten“) zugewiesen und dann der Mittelwert über alle Plattformen gebildet. Die folgenden Plattformen werden berücksichtigt: Windows, Mac, Linux, Android und iOS. Es werden nur Plattformen berücksichtigt die mindestens einen 2%-Marktanteil in Deutschland haben (vgl. Statista, 2021a; Statista, 2021c). Es gilt: Je mehr Plattformen berücksichtigt werden, desto besser ist die Bandbreite an Benutzungskonstellationen und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibungen der zu vergleichenden Prüfverfahren durchgeführt werden.

8.14.2 Verwendete Browser

$$Abdeckung_{\,Browser} = \frac{1}{9} \sum_{browser=1}^{9} x_{browser} \quad mit\ x \in \{0; 0,5; 1\}$$

Für die Metrik „Verwendete Browser“ wird jedem der neun Browser der Wert 0 („in Prüfung nicht enthalten“), 0,5 („in Prüfung optional auf Wunsch des Auftraggebers enthalten“) oder 1 („in Prüfung enthalten“) zugewiesen und dann der Mittelwert über alle Browser gebildet. Die folgenden Browser werden berücksichtigt: Chrome, Firefox, Safari, Edge, Internet Explorer, Opera, iOS Safari, Android Chrome und Samsung Internet. Es werden nur Browser berücksichtigt, die mindestens einen 2%-Marktanteil in Deutschland

haben (vgl. Statista, 2021b; Statista, 2021d). Es gilt: Je mehr Browser berücksichtigt werden, desto besser ist die Bandbreite an Benutzungskonstellationen und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibungen der zu vergleichenden Prüfverfahren durchgeführt werden.

8.14.3 Verwendete Hilfsmittel

$$Abdeckung_{Hilfsmittel} = \frac{1}{6} \sum_{kategorie=1}^{6} Norm(Anzahl_Hilfsmittel_{kategorie})$$

Für die Metrik „Verwendete Hilfsmittel“ werden die Hilfsmittel in die folgenden sechs Kategorien eingeordnet: Screenreader, Bookmarklets, Onlinedienste, Browser- und Betriebssystem-Funktionen, Browser-Erweiterungen sowie sonstige Programme. Für jede Kategorie wird die Anzahl der verwendeten Hilfsmittel normalisiert. Zur Normalisierung der Werte wird für jede Kategorie als minimaler Wert 0 und als maximaler Wert die von allen zu vergleichenden Prüfverfahren höchste Anzahl der Hilfsmittel dieser Kategorie genommen. Anschließend wird der Mittelwert über alle Kategorien gebildet. Die Kategorien wurden aufgrund einer Analyse der Verfahrensbeschreibungen vom BIK BITV-Test (DIAS GmbH, 2019) und dem BITV-Audit (T-Systems MMS, 2020b) ausgewählt. Es gilt: Je mehr Hilfsmittel verwendet werden, desto besser ist die Bandbreite an geprüften Benutzungskonstellationen und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibungen der zu vergleichenden Prüfverfahren durchgeführt werden.

8.15 Öffentlichkeit des Prüfverfahrens

Aufgrund der Expertenrunde wurde für dieses Kriterium ein Gewicht von 4,07% errechnet. Das Kriterium wird anhand der Metrik „Öffentlichkeit des Prüfverfahrens“ bewertet.

$$Öffentlichkeit = \frac{1}{2} \sum_{aspekt=1}^{2} x_{aspekt} \quad mit\ x \in \{0; 1\}$$

Hierfür werden folgende zwei Aspekte mit dem Wert 0 („falsch“) oder 1 („korrekt“) bewertet:

- Es gibt eine öffentliche, detaillierte Beschreibung des Prüfverfahrens inklusive der Prüfschritte/-kriterien.
- Es gibt eine Diskussionsplattform für das Prüfverfahren.

Anschließend wird der Mittelwert über beide Aspekte gebildet. Es gilt: Je mehr Aspekte mit 1 („korrekt“) bewertet werden, desto öffentlicher ist das Prüfverfahren und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibung der zu vergleichenden Prüfverfahren durchgeführt werden.

8.16 Skalierbarkeit hinsichtlich Prüfer

Aufgrund der Expertenrunde wurde für dieses Kriterium ein Gewicht von 4,04% errechnet. Das Kriterium wird anhand der drei Metriken „Einsatz von Menschen mit Behinderung als Prüfer“, „Einsatz eines Teams von Prüfern mit unterschiedlichen Rollen“ und „Integration von Anfänger in ein Prüfteam“ bewertet.

8.16.1 Einsatz von Menschen mit Behinderung als Prüfer

$$Einsatz_{Menschen\ mit\ Behinderung}$$

$$= \frac{1}{2} \sum_{frage=1}^{2} Norm\left(\frac{1}{I} \sum_{interview=1}^{I} x_{frage,interview}\right)$$

$$mit\ I \in \mathbb{N}\ und\ x \in \{1; 2; ...; 6; 7\}$$

Für die Metrik „Einsatz von Menschen mit Behinderung als Prüfer“ werden folgende Interviewfragen betrachtet: „Können Menschen mit Behinderung als Prüfer eingesetzt werden? Hinweis: Als einzelner Prüfer, der alle Prüfschritte durchführt“ und „Können Menschen mit Behinderung als Prüfer eingesetzt werden? Hinweis: Als Prüfer im Team, der bestimmte Prüfschritte durchführt“. Diese Fragen werden anhand einer Likert-Skala von 1 („überhaupt nicht“) bis 7 („vollständig“) bewertet und pro Frage ein Mittelwert über alle Interviews gebildet. Anschließend werden die Mittelwerte jeweils normalisiert. Zur Normalisierung der Werte wird als minimaler Wert 1 und als maximaler Wert 7 genommen. Schließlich wird ein Mittelwert über beide Fragen gebildet. Es gilt: Je höher die Wertung

der Interviewpartner ist, desto besser können Menschen mit Behinderung als Prüfer eingesetzt werden, desto besser ist die Skalierbarkeit und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand von Interviewdaten durchgeführt werden.

8.16.2 Einsatz eines Teams von Prüfern mit unterschiedlichen Rollen

$$Einsatz_{Teams}$$

$$= Norm\left(\frac{1}{I}\sum_{interview=1}^{I} x_{interview}\right) \quad mit\ I \in \mathbb{N}\ und\ x \in \{1;2;...;6;7\}$$

Für die Metrik „Einsatz eines Teams von Prüfern mit unterschiedlichen Rollen“ wird folgende Interviewfrage betrachtet: „Kann ein Team von Prüfern mit unterschiedlichen Rollen eingesetzt werden? Hinweis: Beispiele für Rollen sind Prüfer für unterschiedliche Benutzergruppen, Prüfer mit unterschiedlichen Qualifizierungsleveln oder Schwerpunkten sowie Barrierefreiheitsprüfer und Qualitätssicherung.“. Diese Frage wird anhand einer Likert-Skala von 1 („überhaupt nicht“) bis 7 („vollständig“) bewertet und ein Mittelwert über alle Interviews gebildet. Anschließend wird dieser Mittelwert normalisiert. Zur Normalisierung der Werte wird als minimaler Wert 1 und als maximaler Wert 7 genommen. Es gilt: Je höher die Wertung der Interviewpartner ist, desto besser können Teams von Prüfern mit unterschiedlichen Rollen eingesetzt werden, desto besser ist die Skalierbarkeit und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand von Interviewdaten durchgeführt werden.

8.16.3 Integration von Anfänger in ein Prüfteam

$$Integration_{Anfänger}$$

$$= Norm\left(\frac{1}{I}\sum_{interview=1}^{I} x_{interview}\right) \quad mit\ I \in \mathbb{N}\ und\ x \in \{1;2;...;6;7\}$$

Für die Metrik „Integration von Anfänger in ein Prüfteam“ wird folgende Interviewfrage betrachtet: „Können Anfänger in ein Prüfteam integriert werden?“. Diese Frage wird anhand einer Likert-Skala von 1 („überhaupt nicht“) bis 7 („vollständig“) bewertet und ein Mittelwert über alle Interviews gebildet. Anschließend wird dieser Mittelwert normalisiert. Zur Normalisierung der Werte wird als minimaler Wert 1 und als maximaler Wert

7 genommen. Es gilt: Je höher die Wertung der Interviewpartner ist, desto besser können Anfänger in ein Prüfteam eingesetzt werden, desto besser ist die Skalierbarkeit und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand von Interviewdaten durchgeführt werden.

8.17 Grad der Toolunterstützung

Aufgrund der Expertenrunde wurde für dieses Kriterium ein Gewicht von 3,87% errechnet. Das Kriterium wird anhand der drei Metriken „Multiuserfähigkeit“, „Plattformunabhängigkeit“ und „Automatisierung von Abläufen“ bewertet.

8.17.1 Multiuserfähigkeit

$$Multiuserfähigkeit = \frac{1}{3} \sum_{aspekt=1}^{3} x_{aspekt} \quad mit\ x \in \{0; 1\}$$

Für die Metrik „Multiuserfähigkeit“ werden folgende drei Aspekte mit dem Wert 0 („falsch“) oder 1 („korrekt“) bewertet:

- Ein Wechsel der Prüfer während der Prüfungsdurchführung ist vom Prüfverfahren angedacht.
- Ein Wechsel der Prüfer während der Prüfungsdurchführung ist technisch einfach möglich.
- Die Informationen der Prüfung werden zentral gespeichert.

Anschließend wird der Mittelwert über alle Aspekte gebildet. Es gilt: Je mehr Aspekte mit 1 („korrekt“) bewertet werden, desto besser ist der Grad der Multiuserfähigkeit, desto besser ist der Grad der Toolunterstützung und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibungen der zu vergleichenden Prüfverfahren sowie einer Inspektion der Tools durchgeführt werden.

8.17.2 Plattformunabhängigkeit

$$Plattformunabhängigkeit = x \quad mit\, x \in \{0; 0,34; 0,67; 1\}$$

Die „Plattformunabhängigkeit“ kann wie folgt bewertet werden:

- 1 („komplett unabhängig von Plattformen wie z.B. browserbasierte Anwendung“)
- 0,67 („auf zwei der drei am meisten genutzten Plattformen lauffähig wie z.B. Verwenden von Microsoft Office Produkte“)
- 0,34 („auf einer der drei am meisten genutzten Plattformen lauffähig“)
- 0 („auf einer Plattform unter festgelegten Rahmenbedingungen lauffähig“)

Es gilt: Je mehr Plattformen für das Tool bzw. die Tools geeignet sind, desto besser ist der Grad der Plattformunabhängigkeit, desto besser ist der Grad der Toolunterstützung und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibungen der zu vergleichenden Prüfverfahren sowie einer Inspektion der Tools durchgeführt werden.

8.17.3 Automatisierung von Abläufen

$$Automatisierung = \frac{1}{5}\left(Norm(\frac{1}{I}\sum_{interview=1}^{I} x_{interview}) + \sum_{aspekt=1}^{4} y_{aspekt}\right)$$

$$mit\ \ I \in \mathbb{N}\, und\, x \in \{1; 2; ...; 6; 7\}\, und\, y \in \{0; 1\}$$

Für die Metrik „Automatisierung von Abläufen“ wird folgende Interviewfrage betrachtet: „Inwieweit werden bereits Abläufe der Prüfung automatisiert?“. Diese Frage wird anhand einer Likert-Skala von 1 („überhaupt nicht“) bis 7 („vollständig“) bewertet und ein Mittelwert über alle Interviews gebildet. Anschließend wird dieser Mittelwert normalisiert. Zur Normalisierung der Werte wird als minimaler Wert 1 und als maximaler Wert 7 genommen. Zusätzlich werden folgende Aspekte mit dem Wert 0 („falsch“) oder 1 („korrekt“) bewertet:

- Das Tool unterstützt den Prüfer bei der Webseitenauswahl (Stichprobe).
- Das Tool passt die Prüfschritte/-kriterien je nach Prüfungsumfang und Prüfgegenstand automatisch an.

- Das Tool wird automatisch mit Informationen zur Barrierefreiheit des Prüfgegenstands befüllt.
- Das Tool unterstützt den Prüfer beim Editieren der dokumentierten Probleme.

Der Mittelwert dieser fünf Werte ergibt den Wert der Metrik. Die Aspekte wurden aufgrund einer Analyse der Verfahrensbeschreibungen (DIAS GmbH, 2019; T-Systems MMS, 2020b) sowie Inspektionen der Tools vom BIK BITV-Test (DIAS GmbH, o.D.d) und dem BITV-Audit (T-Systems MMS, o.D.c; T-Systems MMS, o.D.e) ausgewählt. Es gilt: Je mehr Aspekte mit 1 („korrekt“) bewertet werden und je höher die Wertung der Interviewpartner, desto besser ist der Grad der Automatisierung von Abläufen, desto besser ist der Grad der Toolunterstützung und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibungen der zu vergleichenden Prüfverfahren sowie einer Inspektion der Tools durchgeführt werden.

8.18 Lizenz

Aufgrund der Expertenrunde wurde für dieses Kriterium ein Gewicht von 2,96% errechnet. Das Kriterium wird anhand der Metrik „Lizenzbedingungen“ bewertet.

$$Lizenzbedingungen = x \quad mit\, x \in \{0; 0,5; 1\}$$

Die Metrik „Lizenzbedingungen“ kann wie folgt bewertet werden:

- 1 („Prüfverfahren darf ohne besondere Genehmigung verwendet werden“)
- 0,5 („Prüfverfahren darf nur mit besonderer Genehmigung verwendet werden“)
- 0 („Prüfverfahren darf nicht verwendet werden“)

Die Bewertungsoptionen wurden aufgrund einer Analyse der Verfahrensbeschreibungen vom BIK BITV-Test (DIAS GmbH, 2019) und dem BITV-Audit (T-Systems MMS, 2020b) ausgewählt. Es gilt: Je besser die Lizenzbedingungen, desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibungen der zu vergleichenden Prüfverfahren durchgeführt werden.

8.19 Zertifikat

Aufgrund der Expertenrunde wurde für dieses Kriterium ein Gewicht von 2,90% errechnet. Das Kriterium wird anhand der beiden Metriken „Vorhandensein eines Prüfzeichens“ und „Veröffentlichung der Prüfergebnisse“ bewertet.

8.19.1 Vorhandensein eines Prüfzeichens

$$Vorhandensein_{Prüfzeichen} = \frac{1}{2} \sum_{aspekt=1}^{2} x_{aspekt} \quad mit\, x \in \{0;1\}$$

Für die Metrik „Vorhandensein eines Prüfzeichens“ werden folgende beide Aspekte mit dem Wert 0 („falsch“) oder 1 („korrekt“) bewertet:

- Im Prüfbericht ist ein Prüfzeichen enthalten.
- Das Prüfzeichen kann vom Auftraggeber auf seiner Website eingebunden werden.

Anschließend wird der Mittelwert über beide Aspekte gebildet. Es gilt: Je mehr Aspekte mit 1 („korrekt“) bewertet werden, desto besser ist das Zertifikat und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibungen der zu vergleichenden Prüfverfahren durchgeführt werden.

8.19.2 Veröffentlichung der Prüfergebnisse

$$Veröffentlichung_{Prüfergebnisse} = \frac{1}{2} \sum_{aspekt=1}^{2} x_{aspekt} \quad mit\, x \in \{0;1\}$$

Für die Metrik „Veröffentlichung der Prüfergebnisse“ werden folgende zwei Aspekte mit dem Wert 0 („falsch“) oder 1 („korrekt“) bewertet:

- Das Prüfergebnis kann auf Wunsch veröffentlicht werden, sodass der Auftraggeber mit einem Link darauf verweisen kann.
- Das Prüfergebnis darf durch den Auftraggeber veröffentlicht werden.

Anschließend wird der Mittelwert über beide Aspekte gebildet. Die oben genannten Aspekte wurden aufgrund einer Analyse der Verfahrensbeschreibungen vom BIK BITV-Test (DIAS GmbH, 2019) und dem BITV-Audit (T-Systems MMS, 2020b) ausgewählt. Es gilt: Je mehr Aspekte mit 1 („korrekt“) bewertet werden, desto besser ist das Zertifikat und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibungen der zu vergleichenden Prüfverfahren durchgeführt werden.

8.20 Potential für Automatisierung der Überprüfung

Aufgrund der Expertenrunde wurde für dieses Kriterium ein Gewicht von 2,90% errechnet. Nach einer umfassenden Analyse der Verfahrensbeschreibungen des BIK BITV-Tests (DIAS GmbH, 2019) sowie des BITV-Audits (T-Systems MMS, 2020b) und grundlegender Recherche zum Thema Automatisierung mittels künstlicher Intelligenz (KI) lautet die derzeitige Einschätzung, dass die Einflussfaktoren dieses Kriteriums zu komplex sind, um eine valide Aussage treffen zu können und eine vertrauenswürdige Quantifizierung zu finden. Daher erhalten alle Prüfverfahren den Wert 0.

In die Entscheidungsfindung sind folgende Aspekte eingeflossen. Derzeitige Prüftools können eine Prüfung auf Vorhandensein von relevanten Elementen sicherstellen, z.B. Prüfung des Vorhandenseins von Alt-Attributen bei Bildern. Eine kontextsensitive Prüfung und Beurteilung der Sinnhaftigkeit ist derzeit jedoch noch nicht möglich. Relevante Fragen im Bereich der Alternativtexte für Bilder wären zum Beispiel: „Handelt es sich um ein dekoratives Bild, das korrekterweise keinen Alternativtext benötigt?“ oder „Beschreibt der vorhandene Alternativtext das Bild aussagekräftig?“. Zudem können die meisten Prüftools zum jetzigen Zeitpunkt nur Prüfungen auf einzelnen Webseiten durchführen. Prüfschritte/-kriterien, die sich jedoch auf eine ganze Website beziehen oder auf einen seitenübergreifenden Prozess, stellen eine viel größere Herausforderung dar. Zudem müsste die Automatisierung nicht nur das Erkennen eines Problems ermöglichen, sondern auch sinnvolle Problembeschreibungen formulieren aus der ein Leser die Ursache bzw. die Lokalisierung des Problems nachvollziehen kann. Außerdem können manuelle Prüfverfahren die Probleme nach einem Stufenmodell einschätzen. Beispielsweise gibt es beim BITV-Audit folgende drei Stufen: „Leichte Zugänglichkeitseinschränkung“, „Zugänglichkeitshürde“ und „Zugänglichkeitsblockade“ (vgl. T-Systems MMS, 2021a). Diese Einschätzung ist je nach Stufenmodell unterschiedlich, aber in jedem Fall nicht trivial. Hier versuchen Prüfer den Bewertungsspielraum durch einen gemeinsamen Austausch und der Orientierung an vorherigen Prüfergebnissen abzugleichen, um eine homogene Bewertung von Problemen sicher-

zustellen. Auch eine Empfehlung zur Problemlösung ist in vielen Fällen nicht generisch, sondern abhängig von der speziellen Situation.

Außerdem lässt sich keine allgemeingültige Aussage treffen, da die Automatisierung stark von dem verwendeten KI-System abhängt. Es gibt derzeit noch Hürden, die überwunden werden müssen, bevor eine Automatisierung der Prüfschritte vertrauenswürdige Ergebnisse liefert. Beispielsweise kann auf geschützte Bereiche (durch einen Login) noch immer nicht ohne Probleme zugegriffen werden. Je nach Umsetzung des Logins scheitern die intelligenten Systeme heute noch an dieser Hürde, die von einem menschlichen Prüfer ohne Probleme gemeistert wird.

8.21 Zukunftssicherheit für WCAG 3.0

Aufgrund der Expertenrunde wurde für dieses Kriterium ein Gewicht von 2,80% errechnet. Bei der Recherchearbeit hat sich gezeigt, dass sich die WCAG 3.0 noch stark in der Entwicklung befindet. Die Veröffentlichung eines ersten Arbeitsstands ist für Ende 2021 geplant. Die finale Version ist erst für Ende 2022 angekündigt (vgl. W3C, 2021). Auf Grundlage des momentanen Stands kann keine valide Aussage über die Zukunftssicherheit der Prüfverfahren getroffen werden. Daher erhalten alle Prüfverfahren den Wert 0.

Im Folgenden werden einige der bisher bekannten Informationen zur WCAG 3.0 zusammengefasst, um dem Leser einen ersten Eindruck davon zu verschaffen. Es wird jedoch nochmals darauf hingewiesen, dass diese Informationen nicht verbindlich sind und sich noch ändern können. Die WCAG 3.0 ist nicht mit den vorherigen Versionen der WCAG 2.X kompatibel. Zudem soll die WCAG 3.0 auch Inhalte aus den User Agent Accessibility Guidelines[2] 2.0 sowie den Authoring Tool Accessibility Guidelines[3] 2.0 mitaufnehmen. Es wird auch überlegt, wie die WCAG 3.0 die Stichprobenauswahl gestalten kann, um den Umfang der zu prüfenden Webseiten weiter einzugrenzen. Außerdem soll es die Konformitätsstufen A, AA und AAA nicht mehr geben (vgl. W3C, 2021). Derzeit geben die WCAG eine Einschätzung in welchem Maße eine Website rechtliche Richtlinien abdeckt. Eine Aussage der tatsächlichen Barrierefreiheit ist schwer abzuleiten. Beispielsweise gilt eine Website mit nur einem nicht erfüllten Erfolgskriterium der Konformitätsstufe A als nichtkonform zu WCAG 2.0 A, obwohl alle restlichen Erfolgskriterien der Konformitätsstufen A und AA erfüllt sind. Daher wird an einer neuen Metrik zur Messung der Konformität gearbeitet. Diese soll aus folgenden drei Konformitätsstufen bestehen: Bronze (niedrigste),

[2]Richtlinien für Barrierefreiheit der User Agents wie z.B. Browser (vgl. W3C, 2015b)
[3]Richtlinien für Barrierefreiheit der Autorenwerkzeuge (vgl. W3C, 2015a)

Silber (mittlere) und Gold (höchste) (vgl. Holder, 2020). Um die Konformitätsstufe Silber oder Gold erreichen zu können, müssen zusätzlich zu den Konformitätsprüfungen auch holistische Tests durchgeführt werden. Damit sind Prüfungen mit assistiven Technologien gemeint, aber auch Usability-Tests durch Experten und Benutzer (vgl. W3C, 2021).

8.22 Einfachheit der Prüfungsdurchführung

Aufgrund der Expertenrunde wurde für dieses Kriterium ein Gewicht von 2,44% errechnet. Das Kriterium wird anhand der beiden Metriken „Qualifizierungsaufwand für Prüfer“ und „Flüchtigkeitsfehler in Prüfberichten“ bewertet.

8.22.1 Qualifizierungsaufwand für Prüfer

$$Qualifizierungsaufwand$$

$$= Norm_{rev}\left(\frac{1}{I}\sum_{interview=1}^{I} x_{interview}\right) \quad mit\ I \in \mathbb{N}\ und\ x \in \{1;2;...;6;7\}$$

Für die Metrik „Qualifizierungsaufwand für Prüfer“ wird folgende Interviewfrage betrachtet: „Wie hoch ist der Qualifizierungsaufwand für Prüfer?“. Diese Frage wird anhand einer Likert-Skala von 1 („sehr gering“) bis 7 („sehr hoch“) bewertet und ein Mittelwert über alle Interviews gebildet. Anschließend wird dieser Mittelwert normalisiert. Zur Normalisierung der Werte wird als minimaler Wert 1 und als maximaler Wert 7 genommen. Es gilt: Je niedriger die Wertung der Interviewpartner ist, desto geringer ist der Qualifizierungsaufwand, desto einfacher ist die Prüfungsdurchführung und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand von Interviewdaten durchgeführt werden.

8.22.2 Flüchtigkeitsfehler in Prüfberichten

$$Flüchtigkeitsfehler = \frac{1}{W}\sum_{website=1}^{W}\frac{FF_{website}}{KP_{website}} \quad mit\ W \in \mathbb{N}$$

Für die Metrik „Flüchtigkeitsfehler in Prüfberichten“ wird zuerst für jede Website die Anzahl der Flüchtigkeitsfehler ermittelt und dann durch die Anzahl der korrekt positiven Ergebnisse (Bewertung nach WCAG 2.1 AA) geteilt. Bei einem Flüchtigkeitsfehler (FF) handelt es sich um ein Problem, das vom Prüfverfahren gefunden werden müsste, aber

fälschlicherweise nicht gefunden wird. Bei einem korrekt positiven Ergebnis (KP) handelt es sich um ein Problem, das vom Prüfverfahren korrekterweise als Problem identifiziert wird, da es nach WCAG 2.1 AA ein Problem darstellt. Anschließend wird der Mittelwert über alle Websites gebildet. Es gilt: Je weniger Flüchtigkeitsfehler enthalten sind, desto einfacher ist die Prüfungsdurchführung und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Prüfberichtdaten der zu vergleichenden Prüfverfahren durchgeführt werden.

8.23 Organisatorische Anforderungen

Aufgrund der Expertenrunde wurde für dieses Kriterium ein Gewicht von 1,51% errechnet. Das Kriterium wird anhand der beiden Metriken „Finanzieller Aufwand zur Qualifikation einer Organisation“ und „Zeitlicher Aufwand zur Qualifikation einer Organisation“ bewertet.

8.23.1 Finanzieller Aufwand zur Qualifikation einer Organisation

$$Aufwand_{finanziell} = \frac{1}{3}(Norm_{rev}(Qualifizierungskosten)+$$

$$Norm_{rev}(Grundkosten) + Norm_{rev}(Betriebskosten))$$

Für die Metrik „Finanzieller Aufwand zur Qualifikation einer Organisation“ wird zuerst überprüft, ob es bei allen Prüfverfahren möglich ist sich als Organisation zu qualifizieren. Gibt es keine Möglichkeit sich bei einem Prüfverfahren zu qualifizieren, erhält dieses Prüfverfahren den Wert 0. Die anderen Prüfverfahren erhalten den Wert 1. Im Falle, dass eine Qualifikation als Organisation bei allen Prüfverfahren möglich ist, werden die Qualifizierungskosten, Grundkosten sowie die Betriebskosten normalisiert und der Mittelwert gebildet. Die hier beschriebene Metrik wurde aufgrund einer Analyse der Verfahrensbeschreibungen vom BIK BITV-Test (DIAS GmbH, 2019) und dem BITV-Audit (T-Systems MMS, 2020b) erstellt. Es gilt: Je weniger finanzieller Aufwand zur Qualifikation einer Organisation benötigt wird, desto geringer sind die Anforderungen an eine Organisation und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibungen der zu vergleichenden Prüfverfahren durchgeführt werden.

8.23.2 Zeitlicher Aufwand zur Qualifikation einer Organisation

$$Aufwand_{zeitlich} = Norm_{rev}(Benötigte_Zeit)$$

Für die Metrik „Zeitlicher Aufwand zur Qualifikation einer Organisation“ wird zuerst überprüft, ob es bei den Prüfverfahren möglich ist sich als Organisation zu qualifizieren. Gibt es keine Möglichkeit sich bei einem Prüfverfahren zu qualifizieren, erhält dieses Prüfverfahren den Wert 0. Die anderen Prüfverfahren erhalten den Wert 1. Im Falle, dass eine Qualifikation als Organisation bei allen Prüfverfahren möglich ist, wird die benötigte Zeit zur Qualifikation normalisiert. Die hier beschriebene Metrik wurde aufgrund einer Analyse der Verfahrensbeschreibungen vom BIK BITV-Test (DIAS GmbH, 2019) und dem BITV-Audit (T-Systems MMS, 2020b) erstellt. Es gilt: Je kürzer die benötigte Zeit zur Qualifikation einer Organisation ist, desto geringer sind die Anforderungen an eine Organisation und desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibungen der zu vergleichenden Prüfverfahren durchgeführt werden.

8.24 Formate des Prüfberichts

Aufgrund der Expertenrunde wurde für dieses Kriterium ein Gewicht von 1,48% errechnet. Das Kriterium wird anhand der Metrik „Anzahl der Prüfberichtformate“ bewertet.

$$Abdeckung_{Dateiformate} = \frac{1}{4} \sum_{dateiformat=1}^{4} x_{dateiformat} \quad mit\ x \in \{0;1\}$$

Es wird jedem der vier Dateiformate der Wert 0 („nicht abgedeckt“) oder 1 („abgedeckt“) zugewiesen und dann der Mittelwert über alle Dateiformate gebildet. Die folgenden Dateiformate werden berücksichtigt: PDF-Dokument, Textverarbeitungsdokument, Tabellenkalkulationsdokument sowie HTML/Webseite. Sie wurden durch eine Recherche der gängigen Dateiformate von Berichten identifiziert und in die genannten Typen gruppiert. Es gilt: Je mehr Dateiformate vom Prüfverfahren abgedeckt werden, desto höher ist der normalisierte Wert dieser Metrik. Die Auswertung kann anhand der Verfahrensbeschreibungen der zu vergleichenden Prüfverfahren durchgeführt werden.

9 Vergleich der Prüfverfahren anhand des erstellten Kriterienkatalogs

Auf Basis des erstellten Kriterienkatalogs wird der Vergleich der ausgewählten Prüfverfahren durchgeführt. In Abschnitt 9.1 wird das Vorgehen des Vergleichs beschrieben. In Abschnitt 9.2 werden die aus dem Vergleich resultierenden Ergebnisse zusammengefasst und anschließend in Abschnitt 9.3 im Kontext der verwandten Arbeiten und den Zielen dieser Arbeit diskutiert.

9.1 Vorgehensbeschreibung

Der Vergleich der Prüfverfahren beruht auf den derzeit gültigen Verfahrensbeschreibungen sowie den eigens für den Vergleich erstellten Prüfberichten. Die erstellten Prüfberichte wurden in einem Workshop mit Barrierefreiheitsexperten aufbereitet. Zudem wurde bei der Bewertung einzelner Kriterien die subjektive Bewertung von Barrierefreiheitsexperten berücksichtigt, hierfür wurden Interviews durchgeführt.

9.1.1 Erstellung der Prüfberichte

In einem ersten Schritt wurden die Websites ausgewählt, die mit beiden Prüfverfahren geprüft werden sollten: die Internetpräsenz der Hochschule der Medien[1](HdM) sowie die Internetpräsenz der Bundesfachstelle für Barrierefreiheit[2](BfB). Vorab wurde hierfür schriftlich das Einverständnis der BfB eingeholt. Diese beiden Websites wurden unter der Annahme ausgewählt, dass sie sich in dem Grad ihrer Barrierefreiheit unterscheiden. So können die Ergebnisse verglichen werden, wenn die beiden Prüfverfahren auf Basis einer weniger barrierefreien und einer eher barrierefreien Website durchgeführt werden.

[1] erreichbar unter `https://www.hdm-stuttgart.de/`

[2] erreichbar unter `https://www.bundesfachstelle-barrierefreiheit.de/DE/Home/home_node.html`

Für eine Prüfung, die zu einer offiziellen Zertifizierung führen soll, muss bei beiden Prüfverfahren eine repräsentative Stichprobe der Website ausgewählt werden. Dies war aufgrund des damit einhergehenden, höheren Aufwands im Rahmen dieser Masterarbeit nicht möglich. Es wurde entschieden, dass nur die beiden Startseiten der Internetpräsenzen geprüft werden, da sie eine Auswahl an Layouts, Navigationsmenüs sowie weiteren Elementen beinhalten und so viele Prüfschritte anwendbar sind.

Die Prüfungen wurden von einem Mitarbeiter der T-Systems MMS durchgeführt, weil er für beide Prüfverfahren offiziell qualifiziert ist. Jede Website wurde sowohl mit dem BIK BITV-Test als auch dem BITV-Audit geprüft. So stehen vier Prüfberichte desselben Prüfers zur Verfügung, diese sind im elektronischen Anhang zu finden. Zudem wurde die zur Durchführung der Prüfung benötigte Zeit gemessen und dokumentiert. Die Zeitmessung umfasst alle Prozessschritte der Prüfung (Vorbereitungen für die Prüfung, Durchführung der Prüfung, Qualitätssicherung, Prüfberichterstellung). Soweit dies technisch möglich war, wurde der Zustand der beiden Webseiten zu verschiedenen Zeitpunkten des Prüfprozesses gespeichert, um mögliche Abweichungen zu einem späteren Zeitpunkt nachvollziehen zu können. Die Dateien sind im elektronischen Anhang zu finden.

Bis auf die Stichprobenauswahl wurde das Prüfverfahren nach dem offiziellen Vorgehen des abschließenden BIK BITV-Tests durchgeführt. Als Zweitprüfer wurde ein qualifizierter Prüfer des BIK-Prüfverbunds eingesetzt. Die Qualitätssicherung wurde durch einen dafür qualifizierten Prüfer des BIK-Prüfverbunds durchgeführt. Zur Durchführung der Prüfung und Dokumentation der gefundenen Probleme wurde das offizielle BITV-Test-Prüftool verwendet. Da der Zeitraum der Prüfung durch äußere Einflüsse wie weiteren Projektarbeiten und Krankheit länger war als normalerweise üblich, gab es wenige Änderungen an den Webseiten zwischen der Prüfung des Erstprüfers und des Zweitprüfers sowie der Qualitätssicherung. Diese Abweichungen konnten durch die Sicherungen der Webseitenzustände problemlos nachvollzogen werden.

Bis auf die Stichprobenauswahl wurde das Prüfverfahren nach offiziellem Vorgehen des BITV-Audit durchgeführt. Ein fester Bestandteil des Audits ist die Auswahl der anzuwendenden Prüfkriterien. Um die Prüfungen der beiden Prüfverfahren möglichst vergleichbar zu gestalten, wurde die Auswahl auf die Prüfkriterien beschränkt, die den WCAG 2.1 AA zugeordnet sind. Das Teamreview sowie das Abnahmereview wurden durch eine qualifizierte Mitarbeiterin der T-Systems MMS durchgeführt.

Um Verzerrungseffekte zu vermeiden, wurde auf folgende Aspekte geachtet: Die Prüfverfahren wurden in abwechselnder Reihenfolge angewendet. Daher wurde die Startseite der HdM zuerst mit dem BITV-Audit und anschließend mit dem BIK BITV-Test geprüft.

	BIK BITV-Test	**BITV-Audit**
HdM	*Seite 1:* 33 von 50 BITV-Anforderungen erfüllt *Seite 1:* nicht BITV-konform *Gesamt:* 0 von 1 der für die Prüfung ausgewählten Seiten wurden als BITV-konform bewertet	*Gesamt:* 31 von 50 Prüfkriterien erfüllt *Gesamt:* nicht BITV-konform *Zugänglichkeitsaussage:* sehbeeinträchtigt: eingeschränkt blind: ungenügend motorisch beeinträchtigt: mangelhaft hörgeschädigt: sehr gut kognitiv beeinträchtigt: eingeschränkt
BfB	*Seite 1:* 41 von 50 BITV-Anforderungen erfüllt *Seite 1:* nicht BITV-konform *Gesamt:* 0 von 1 der für die Prüfung ausgewählten Seiten wurden als BITV-konform bewertet	*Gesamt:* 36 von 50 Prüfkriterien erfüllt *Gesamt:* teilweise BITV-konform *Zugänglichkeitsaussage:* sehbeeinträchtigt: eingeschränkt blind: eingeschränkt motorisch beeinträchtigt: gut hörgeschädigt: sehr gut kognitiv beeinträchtigt: eingeschränkt

Tabelle 9.1: Überblick über die Prüfberichte der HdM und BfB

Anschließend wurde die Startseite der BfB zuerst mit dem BIK BITV-Test und dann mit dem BITV-Audit geprüft. Außerdem wurden alle beteiligten Experten darauf hingewiesen, persönlich bevorzugte Gruppierungen von Prüfschritten oder individuelle Vorgehensweisen zu vermeiden. So wurde sichergestellt, dass die offizielle Vorgehensweise und Reihenfolge der Prüfschritte eingehalten wurde. Die Ergebnisse der Prüfberichte werden in Tabelle 9.1 zusammengefasst.

9.1.2 Aufbereitung der Prüfberichtergebnisse durch Experten

Die Ergebnisse der drei Kriterien „Korrektheit“, „Vollständigkeit“ und „Einfachheit der Prüfungsdurchführung“ basieren auf einer Einordnung der gefundenen Probleme in die folgenden vier Kategorien:

- Falsch positives Ergebnis: Ein Problem, das vom Prüfverfahren fälschlicherweise als Problem identifiziert wird, aber kein Problem darstellt.

- Falsch negatives Ergebnis: Ein Problem, das vom Prüfverfahren fälschlicherweise nicht als Problem identifiziert wird, aber ein Problem darstellt.
- Korrekt positives Ergebnis: Ein Problem, das korrekterweise vom Prüfverfahren als Problem identifiziert wird.
- Flüchtigkeitsfehler: Ein Problem, das vom Prüfverfahren als Problem identifiziert werden würde, aber fälschlicherweise vom Prüfer beim Anwenden des Prüfverfahrens nicht erkannt wird.

Da die Einordnung der Probleme in diese Kategorien nicht trivial ist, wurde die Idee entwickelt, die Klassifikation der Probleme im Rahmen eines Workshops mit Barrierefreiheitsexperten zu erarbeiten. Vorab wurden die Probleme der Prüfberichte für jede Website in einer Liste gesammelt. Dabei wurden Probleme, die von beiden Prüfverfahren identifiziert wurden, in der gleichen Zeile positioniert. Hinweise, die keine Auswirkungen auf die Bewertung haben, wurden nicht in die Liste aufgenommen. In mehreren virtuellen Sitzungen trafen sich vier Barrierefreiheitsexperten: der Erstprüfer der Masterarbeit, die Praxisbetreuerin der Masterarbeit, der für die Prüfberichte verantwortliche Prüfer sowie der Prüfer, der als Zweitprüfer beim BIK BITV-Test beteiligt war.

Das Ziel des Workshops war die Einordnung der gefundenen Probleme in eine der oben definierten vier Kategorien. Vorab wurde als Bewertungsstandard WCAG 2.1 AA festgelegt, da die WCAG 2.1 AA Erfolgskriterien auch bei der Expertengewichtung der Vergleichskriterien das höchste Gewicht erhielten. Zu Beginn des Workshops stellte sich heraus, dass die Einordnung der Probleme zusätzlich auch aus Nutzersicht vorgenommen werden muss, da es bei diesen beiden Sichtweisen zu zahlreichen abweichenden Ergebnissen kommen kann. Die Klassifikation der Probleme der BfB Prüfberichte wurden wie folgt vorgenommen: Jedes Problem wurde gemeinsam diskutiert und wenn nötig anhand der Sicherung der Website nochmals genauer betrachtet. Es wurde überprüft, ob ein dokumentiertes Problem auch tatsächlich laut Verfahrensbeschreibung als Problem bewertet wird. Aber auch, ob ein nicht dokumentiertes Problem vom Prüfverfahren hätte erkannt werden müssen. Anschließend wurde eine gemeinsame Einordnung des Problems anhand von WCAG 2.1 AA und aus Nutzersicht durchgeführt. Die Bewertungen sowie deren Begründungen wurden für alle Experten sichtbar durch die Autorin der Arbeit notiert. Nach der gemeinsamen Aufbereitung der BfB Prüfberichte, bereitete die Autorin die HdM Prüfberichte analog auf. Diese aufbereitete Liste wurde dann von den Experten vor der nächsten gemeinsamen Sitzung geprüft, sodass gezielt unklare Bewertungen und offene Fragen diskutiert werden konnten. Insgesamt umfassten die gemeinsamen Sitzungen zehn Stunden.

	BIK BITV-Test		BITV-Audit	
	HdM	BfB	HdM	BfB
Falsch positive Ergebnisse (WCAG 2.1 AA)	3	2	7	11
Falsch positive Ergebnisse (Nutzersicht)	1	3	0	1
Falsch negative Ergebnisse (WCAG 2.1 AA)	12	5	6	3
Falsch negative Ergebnisse (Nutzersicht)	20	17	9	4
Korrekt positive Ergebnisse (WCAG 2.1 AA)	62	24	68	26
Korrekt positive Ergebnisse (Nutzersicht)	64	23	75	36
Flüchtigkeitsfehler	7	7	12	3

Tabelle 9.2: Überblick über aufbereitete Ergebnisse der Prüfberichte

Eine Übersicht der Aufbereitungsergebnisse ist in Tabelle 9.2 zu sehen. Die genaue Auflistung der eingeordneten Probleme ist im elektronischen Anhang zu finden. Im Folgenden wird von der Expertenwertung gesprochen, wenn Bezug zu den Ergebnissen des Workshops genommen wird.

Folgende Ursachen wurden von den Experten während des Workshops für einige unterschiedliche Ergebnisse der Prüfverfahren identifiziert: Unterschiede bei der Erkennung von Problemen bezüglich der Kontrastanpassung können daraus resultieren, dass der BITV-Audit mit der Kontrastanpassung des Betriebssystems arbeitet und der BIK BITV-Test nicht. Außerdem können Unterschiede bei der Erkennung von Problemen bezüglich dem Textüberlauf davon kommen, dass die beiden Prüfverfahren unterschiedliche Auflösungen bei der Prüfung verwenden. Hier unterscheiden sich die Rahmenbedingungen der Prüfung.

Die Erwartung, dass es sich bei der Website der HdM um eine weniger barrierefreie Website und bei der Website der BfB um eine eher barrierefreie Website handelt, hat sich durch die Prüfberichte und Expertenwertungen bestätigt. Es wurden insgesamt bei der HdM Webseite 74 Probleme nach WCAG 2.1 AA und 84 Probleme aus Nutzersicht identifiziert. Bei der BfB Webseite hingegen wurden nur 29 Probleme nach WCAG 2.1 AA und 40 Probleme aus Nutzersicht festgestellt. Die Anzahl der Probleme setzt sich jeweils aus den falsch negativen und korrekt positiven Ergebnissen zusammen.

In den Ergebnissen der Expertenwertung zeigt sich zudem, dass 10 Probleme der HdM Webseite nicht durch die WCAG 2.1 als Problem bewertet werden würden, aber für Benutzer ein reales Problem darstellen. Bei der Webseite der BfB sind es sogar 13 Probleme, die durch die WCAG 2.1 nicht als Problem erkannt werden würden, aber für Benutzer

ein reales Problem darstellen. Es gibt jedoch bei der BfB Webseite auch 2 Probleme, die durch die WCAG 2.1 als Problem bewertet werden, aber für Benutzer kein reales Problem darstellen.

9.1.3 Interviews

Bei den folgenden vier der insgesamt 22 Kriterien wurde die subjektive Einschätzung von Barrierefreiheitsexperten bei der Bewertung berücksichtigt: „Qualität des Prüfberichts“, „Skalierbarkeit hinsichtlich Prüfer“, „Grad der Toolunterstützung“ und „Einfachheit der Prüfungsdurchführung“. Hierfür wurden von der Autorin der Arbeit halbstrukturierte Experteninterviews im Zeitraum vom 08.03.2021 bis 22.03.2021 virtuell durchgeführt. Die Interviews wurden mit dem Einverständnis der Experten aufgezeichnet. Es haben insgesamt sechs Experten teilgenommen, sodass am Ende nicht nur die Meinung eines Experten ausschlaggebend war. Es wurden drei Mitglieder des BIK-Prüfverbunds zum BIK BITV-Test interviewt und drei Mitarbeiter der T-Systems MMS zum BITV-Audit. Ein Mitarbeiter der T-Systems MMS führte das Interview zusätzlich für den BIK BITV-Test durch, da er auch im BIK-Prüfverbund ist.

Der Interviewfragebogen bestand aus sechs Items[3]. Die Items wurden bereits in Kapitel 8 genannt. Die zugehörige Likert-Skala reichte von 1 („überhaupt nicht“ bzw. „sehr gering“) bis 7 („vollständig“ bzw. „sehr hoch“). Die Interviewpartner wurden gebeten, jedem Item einen Wert auf der Skala zuzuordnen und ihre Entscheidung zu begründen, um so ein tieferes Verständnis für die Bewertung zu erhalten. Um Verzerrungseffekte zu vermeiden, wurden die Items den Interviewpartnern in unterschiedlichen Reihenfolgen vorgelegt. Es wurde vorab eine einfache Ethikprüfung durch den Erstprüfer der Masterarbeit durchgeführt. Die Vorlage der Einverständniserklärung, eine Übersicht der Items, eine Übersicht der Ergebnisse sowie die anonymisierten Notizen der einzelnen Interviews sind im elektronischen Anhang zu finden.

9.2 Ergebnisse

Der BITV-Audit von T-Systems MMS erreicht den höchsten Gesamtscore mit 0,69. Der BIK BITV-Test erreicht einen Gesamtscore von 0,50. Der Gesamtscore entspricht

[3]Ein Item ist eine Frage bzw. eine Aussage in einem Fragebogen (vgl. Hogrefe AG, 2016).

der Summe des Produkts aus Kriteriumscore und dem jeweiligen Gewicht. Eine ausführliche Definition ist in Abschnitt 8.2 zu finden. Die Ergebnisse der 22 Kriterien und ihren jeweiligen Gewichten sind in Tabelle 9.3 zu sehen. Alle Ergebnisse werden auf zwei Nachkommastellen gerundet. Alle Rohdaten sind im elektronischen Anhang zu finden. Bei beiden Prüfverfahren werden die Kriterien „Potential zur Automatisierung der Überprüfung“ und „Zukunftssicherheit für WCAG 3.0“ nicht bewertet aus den in Abschnitt 8.20 und Abschnitt 8.21 genannten Gründen. Daher werden die beiden Kriterien mit 0,00 bewertet. Der maximal erreichbare Gesamtscore beträgt dadurch nur noch 0,94 statt 1,00. Im Nachfolgenden werden diese beiden Kriterien nicht mehr berücksichtigt.

Bei einer Differenz ($Diff$) von weniger als 0,10 zwischen den Kriteriumscores der beiden Prüfverfahren ($Diff < 0{,}10$), werden nur geringe Abweichungen beobachtet. Die Prüfverfahren weisen bei diesem Kriterium große Gemeinsamkeiten auf.

Bei einer Differenz ($Diff$) von mindestens 0,50 zwischen den Kriteriumscores der beiden Prüfverfahren ($Diff \geq 0{,}50$), werden starke Abweichungen beobachtet. Die Prüfverfahren weisen bei diesem Kriterium große Unterschiede auf.

Bei den folgenden vier Kriterien haben die Prüfverfahren große Gemeinsamkeiten: „Abdeckung von WCAG 2.1 AA“ ($Diff = 0{,}00$), „Korrektheit“ ($Diff = 0{,}04$), „Grad der Toolunterstützung“ ($Diff = 0{,}08$) sowie „Einfachheit der Prüfungsdurchführung“ ($Diff = 0{,}05$).

Bei den folgenden sieben Kriterien unterscheiden sich die Prüfverfahren stark: „Abdeckung von zusätzlichen Kriterien aus EN 301 549 Tab. A.1“ ($Diff = 0{,}71$; zugunsten BITV-Audit), „Abdeckung von zusätzlichen Kriterien aus BITV 2.0“ ($Diff = 1{,}00$; zugunsten BITV-Audit), „Abdeckung von zusätzlichen Kriterien aus WCAG 2.1 AAA“ ($Diff = 1{,}00$; zugunsten BITV-Audit), „Optionale Inputformate“ ($Diff = 1{,}00$; zugunsten BITV-Audit), „Lizenz“ ($Diff = 0{,}50$; zugunsten BIK BITV-Test), „Öffentlichkeit des Prüfverfahrens“ ($Diff = 1{,}00$; zugunsten BIK BITV-Test) sowie „Organisatorische Anforderungen“ ($Diff = 1{,}00$; zugunsten BIK BITV-Test).

Kriterium	Gewicht (in %)	BIK BITV-Test	BITV-Audit
Abdeckung von WCAG 2.1 AA	7,72	1,00	1,00
Vollständigkeit	7,39	0,75	0,90
Qualität des Prüfberichts	6,67	0,30	0,79
Abdeckung von zusätzlichen Kriterien aus EN 301 549 Tab. A.1	6,64	0,02	0,74
Korrektheit	6,32	0,94	0,90
Qualität der Stichprobe	5,96	0,73	0,93
Qualitätssicherung	5,58	0,73	0,60
Optionale Inputformate	5,57	0,00	1,00
Abdeckung von zusätzlichen Kriterien aus BITV 2.0	5,53	0,00	1,00
Aufwand	4,77	0,17	0,29
Abdeckung von zusätzlichen Kriterien aus WCAG 2.1 AAA	4,70	0,00	1,00
Bandbreite an geprüften Benutzungskonstellationen	4,19	0,63	0,74
Öffentlichkeit des Prüfverfahrens	4,07	1,00	0,00
Skalierbarkeit hinsichtlich Prüfer	4,04	0,63	0,80
Grad der Toolunterstützung	3,87	0,52	0,60
Lizenz	2,96	0,50	0,00
Zertifikat	2,90	1,00	0,75
Potential für Automatisierung der Überprüfung	2,90	0,00 (n.b.)	0,00 (n.b.)
Zukunftssicherheit für WCAG 3.0	2,80	0,00 (n.b.)	0,00 (n.b.)
Einfachheit der Prüfungsdurchführung	2,44	0,54	0,59
Organisatorische Anforderungen	1,51	1,00	0,00
Formate des Prüfberichts	1,48	0,50	0,25
Gesamtscore	**100**	**0,50**	**0,69**

Tabelle 9.3: Ergebnisse aller Kriterien für den BIK BITV-Test und den BITV-Audit

Bei einem Kriteriumscore (KS) von weniger als 0,25 ($KS < 0{,}25$), wird weniger als ein Viertel der möglichen Wertung erreicht. Bei diesem Kriterium liegt eine Schwäche des Prüfverfahrens.

Bei einem Kriteriumscore (KS) von mindestens 0,75 ($KS \geq 0{,}75$), wird mindestens Dreiviertel der möglichen Wertung erreicht. Bei diesem Kriterium liegt eine Stärke des Prüfverfahrens.

Bei den folgenden fünf Kriterien liegen die Schwächen des BIK BITV-Tests: „Abdeckung von zusätzlichen Kriterien aus EN 301 549 Tab. A.1“ ($KS = 0{,}02$), „Optionale Inputformate“ ($KS = 0{,}00$), „Abdeckung von zusätzlichen Kriterien aus BITV 2.0“ ($KS = 0{,}00$), „Aufwand“ ($KS = 0{,}17$) sowie „Abdeckung von zusätzlichen Kriterien aus WCAG 2.1 AAA“ ($KS = 0{,}00$). Bei den folgenden sechs Kriterien liegen die Stärken des Prüfverfahrens: „Abdeckung von WCAG 2.1 AA“ ($KS = 1{,}00$), „Vollständigkeit“ ($KS = 0{,}75$), „Korrektheit“ ($KS = 0{,}94$), „Öffentlichkeit des Prüfverfahrens“ ($KS = 1{,}00$), „Zertifikat“ ($KS = 1{,}00$) und „Organisatorische Anforderungen“ ($KS = 1{,}00$).

Bei den folgenden drei Kriterien liegen die Schwächen des BITV-Audits: „Öffentlichkeit des Prüfverfahrens“ ($KS = 0{,}00$), „Lizenz“ ($KS = 0{,}00$) und „Organisatorische Anforderungen“ ($KS = 0{,}00$). Bei den folgenden zehn Kriterien liegen die Stärken des Prüfverfahrens: „Abdeckung von WCAG 2.1 AA“ ($KS = 1{,}00$), „Vollständigkeit“ ($KS = 0{,}90$), „Qualität des Prüfberichts“ ($KS = 0{,}79$), „Korrektheit“ ($KS = 0{,}90$), „Qualität der Stichprobe“ ($KS = 0{,}93$), „Optionale Inputformate“ ($KS = 1{,}00$), „Abdeckung von zusätzlichen Kriterien aus BITV 2.0“ ($KS = 1{,}00$), „Abdeckung von zusätzlichen Kriterien aus WCAG 2.1 AAA“ ($KS = 1{,}00$), „Skalierbarkeit hinsichtlich Prüfer“ ($KS = 0{,}80$) sowie „Zertifikat“ ($KS = 0{,}75$).

Im nächsten Abschnitt werden die zehn Kriterien mit den höchsten Gewichten (Top-10-Kriterien) betrachtet. Anschließend werden die Kriterien in die folgenden vier thematischen Bereiche gruppiert: Abdeckung verschiedener Standards, Qualität der Prüfung und der Prüfergebnisse, Effizienz der Prüfung und Rahmenbedingungen des Prüfverfahrens. Es werden die Ergebnisse der jeweiligen Kriterien und zugehörigen Metriken genauer betrachtet.

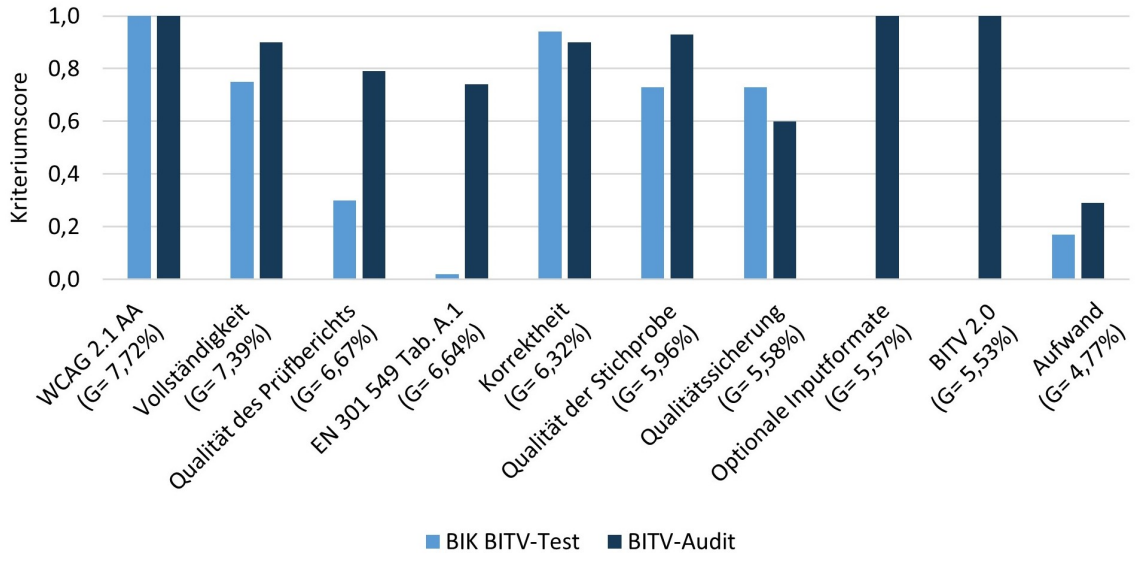


Abbildung 9.1: Ergebnisse der zehn Kriterien mit den höchsten Gewichten

9.2.1 Übersicht der Top-10-Kriterien

Abbildung 9.1 zeigt die Kriteriumscores der zehn Kriterien mit den höchsten Gewichten. Von links nach rechts werden die Gewichte (G) kleiner. Bei den folgenden beiden Kriterien haben die Prüfverfahren große Gemeinsamkeiten: „Abdeckung von WCAG 2.1 AA“ ($Diff = 0{,}00$) und „Korrektheit“ ($Diff = 0{,}04$). Bei den folgenden drei Kriterien unterscheiden sich die Prüfverfahren stark: „Abdeckung von zusätzlichen Kriterien aus EN 301 549 Tab. A.1“ ($Diff = 0{,}71$; zugunsten BITV-Audit), „Optionale Inputformate“ ($Diff = 1{,}00$; zugunsten BITV-Audit) und „Abdeckung von zusätzlichen Kriterien aus BITV 2.0“ ($Diff = 1{,}00$; zugunsten BITV-Audit).

Bei den folgenden vier Top-10-Kriterien liegen die Schwächen des BIK BITV-Tests: „Abdeckung von zusätzlichen Kriterien aus EN 301 549 Tab. A.1“ ($KS = 0{,}02$), „Optionale Inputformate“ ($KS = 0{,}00$), „Abdeckung von zusätzlichen Kriterien aus BITV 2.0“ ($KS = 0{,}00$) sowie „Aufwand“ ($KS = 0{,}17$). Bei den folgenden drei Top-10-Kriterien liegen die Stärken des Prüfverfahrens: „Abdeckung von WCAG 2.1 AA“ ($KS = 1{,}00$), „Vollständigkeit“ ($KS = 0{,}75$) und „Korrektheit“ ($KS = 0{,}94$)

Der BITV-Audit hat keine Schwächen bei den Top-10-Kriterien. Bei den folgenden sieben Top-10-Kriterien liegen die Stärken des Prüfverfahrens: „Abdeckung von WCAG 2.1 AA“ ($KS = 1{,}00$), „Vollständigkeit“ ($KS = 0{,}90$), „Qualität des Prüfberichts“ ($KS = 0{,}79$),

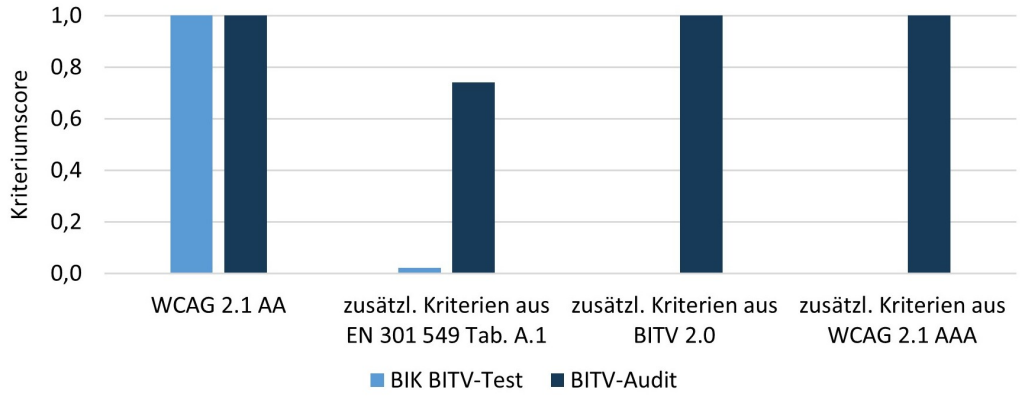


Abbildung 9.2: Ergebnisse der vier Kriterien, die die Abdeckung von Standards betreffen

„Korrektheit“ ($KS = 0{,}90$), „Qualität der Stichprobe“ ($KS = 0{,}93$), „Optionale Inputformate“ ($KS = 1{,}00$) und „Abdeckung von zusätzlichen Kriterien aus BITV 2.0“ ($KS = 1{,}00$).

9.2.2 Abdeckung verschiedener Standards

Die Ergebnisse folgender vier Kriterien werden in diesem Abschnitt zusammengefasst: „Abdeckung von WCAG 2.1 AA“ ($G = 7{,}72\%$), „Abdeckung von zusätzlichen Kriterien aus EN 301 549 Tab. A.1“ ($G = 6{,}64\%$), „Abdeckung von zusätzlichen Kriterien aus BITV 2.0“ ($G = 5{,}53\%$) und „Abdeckung von zusätzlichen Kriterien aus WCAG 2.1 AAA“ ($G = 4{,}70\%$). Die Kriteriumscores der aufgeführten Kriterien sind in Abbildung 9.2 zu sehen.

Alle WCAG 2.1 Erfolgskriterien der Konformitätsstufe A und AA werden von beiden Prüfverfahren ($KS = 1{,}00; Diff = 0{,}00$) vollständig abgedeckt. Bei den drei anderen Kriterien unterscheiden sich die Prüfverfahren stark.

Vom BIK BITV-Test wird nur ein zusätzliches Kriterium der EN 301 549 Tab. A.1 ($KS = 0{,}02$) abgedeckt. Das Prüfverfahren beinhaltet zudem auch keine Prüfschritte, die auf die zusätzlichen Kriterien der BITV 2.0 ($KS = 0{,}00$) abzielen. Zwar bieten einzelne Prüfstellen des BIK-Prüfverbunds die Möglichkeit an, auch die WCAG 2.1 Erfolgskriterien der Konformitätsstufe AAA zu prüfen, dies ist jedoch nicht Teil des offiziellen Prüfverfahrens ($KS = 0{,}00$). Die Grenzen des BIK BITV-Tests werden in einem Artikel

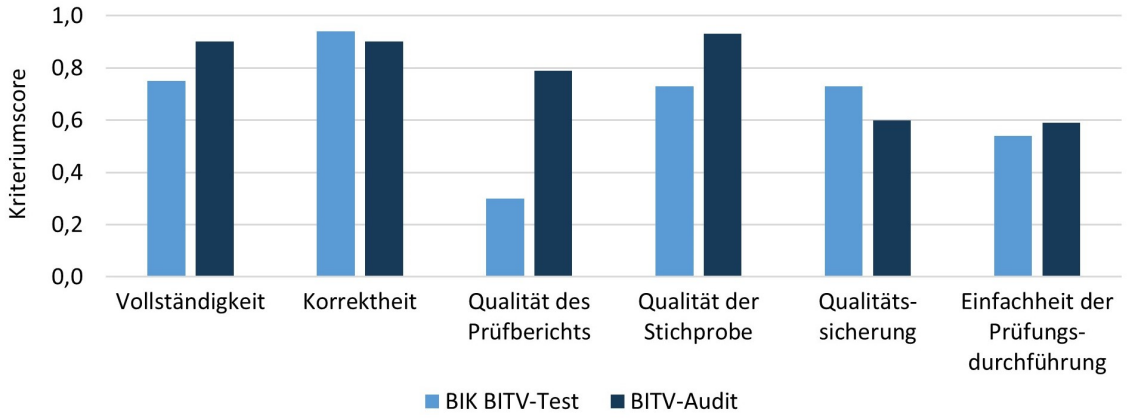


Abbildung 9.3: Ergebnisse der sechs Kriterien, die die Qualität der Prüfung und Prüfergebnisse betreffen

klar aufgezeigt, sodass deutlich wird welche Anforderungen das Prüfverfahren prüft und welche nicht (vgl. DIAS GmbH, o.D.e).

Der BITV-Audit prüft 31 der 42 zusätzlichen Kriterien der EN 301 549 Tab. A.1 ($KS = 0{,}74$). Zudem prüft er alle zusätzlichen Kriterien der BITV 2.0 ($KS = 1{,}00$). Auch die WCAG 2.1 Erfolgskriterien der Konformitätsstufe AAA werden vollständig vom Prüfverfahren abgedeckt ($KS = 1{,}00$). In der Checkliste des BITV-Audits wird die Zuordnung der abgedeckten Kriterien aufgezeigt und so auch die Grenzen dieses Prüfverfahrens deutlich (vgl. T-Systems MMS, o.D.c).

9.2.3 Qualität der Prüfung und Prüfergebnisse

Die Ergebnisse folgender sechs Kriterien werden in diesem Abschnitt zusammengefasst: „Vollständigkeit“ ($G = 7{,}39\%$), „Qualität des Prüfberichts“ ($G = 6{,}67\%$), „Korrektheit“ ($G = 6{,}32\%$), „Qualität der Stichprobe“ ($G = 5{,}96\%$), „Qualitätssicherung“ ($G = 5{,}58\%$) und „Einfachheit der Prüfungsdurchführung“ ($G = 2{,}44\%$). Die Kriteriumscores der aufgeführten Kriterien sind in Abbildung 9.3 zu sehen.

Bei den folgenden beiden Kriterien haben die Prüfverfahren große Gemeinsamkeiten: „Korrektheit“ ($Diff = 0{,}04$) und „Einfachheit der Prüfungsdurchführung“ ($Diff = 0{,}05$). Bei keinem Kriterium unterscheiden sich die Prüfverfahren stark.

Der BIK BITV-Test hat keine Schwächen bei den betrachteten Kriterien. Bei den folgenden zwei Kriterien liegen die Stärken des Prüfverfahrens: „Vollständigkeit“ ($KS = 0{,}75$) und

„Korrektheit“ ($KS = 0{,}94$).

Der BITV-Audit hat keine Schwächen bei den betrachteten Kriterien. Bei den folgenden vier Kriterien liegen die Stärken des Prüfverfahrens: „Vollständigkeit“ ($KS = 0,90$), „Korrektheit“ ($KS = 0,90$), „Qualität des Prüfberichts“ ($KS = 0,79$) sowie „Qualität der Stichprobe“ ($KS = 0,93$). Die folgenden Abschnitte gehen genauer auf die einzelnen Kriterien ein.

Vollständigkeit und Korrektheit

Abbildung 9.4 zeigt eine Übersicht der beiden Kriterien „Vollständigkeit“ ($G = 7{,}39\%$) und „Korrektheit“ ($G = 6{,}32\%$) und den zugehörigen Metriken. Die Werte der Metriken „Bewertung nach WCAG 2.1“ und „Bewertung aus Nutzersicht“ beruhen auf der in Unterabschnitt 9.1.2 beschriebenen Expertenwertung.

> Bei einer Differenz ($Diff$) von weniger als 0,10 zwischen den Metrikscores der beiden Prüfverfahren ($Diff < 0{,}10$), werden nur geringe Abweichungen beobachtet. Die Prüfverfahren weisen bei dieser Metrik große Gemeinsamkeiten auf.
>
> Bei einer Differenz ($Diff$) von mindestens 0,50 zwischen den Metrikscores der beiden Prüfverfahren ($Diff \geq 0{,}50$), werden starke Abweichungen beobachtet. Die Prüfverfahren weisen bei dieser Metrik große Unterschiede auf.

Der Kriteriumscore für das Kriterium „Vollständigkeit“ beträgt für den BIK BITV-Test 0,75 und für den BITV-Audit 0,90. Bei der Metrik „Bewertung nach WCAG 2.1 AA“ ($Diff = 0{,}08$) haben die Prüfverfahren große Gemeinsamkeiten. Bei keiner Metrik unterscheiden sich die Prüfverfahren stark.

> Bei einem Metrikscore (MS) von weniger als 0,25 ($MS < 0{,}25$), wird weniger als ein Viertel der möglichen Wertung erreicht. Bei dieser Metrik liegt eine Schwäche des Prüfverfahrens.
>
> Bei einem Metrikscore (MS) von mindestens 0,75 ($MS \geq 0{,}75$), wird mindestens Dreiviertel der möglichen Wertung erreicht. Bei dieser Metrik liegt eine Stärke des Prüfverfahrens.

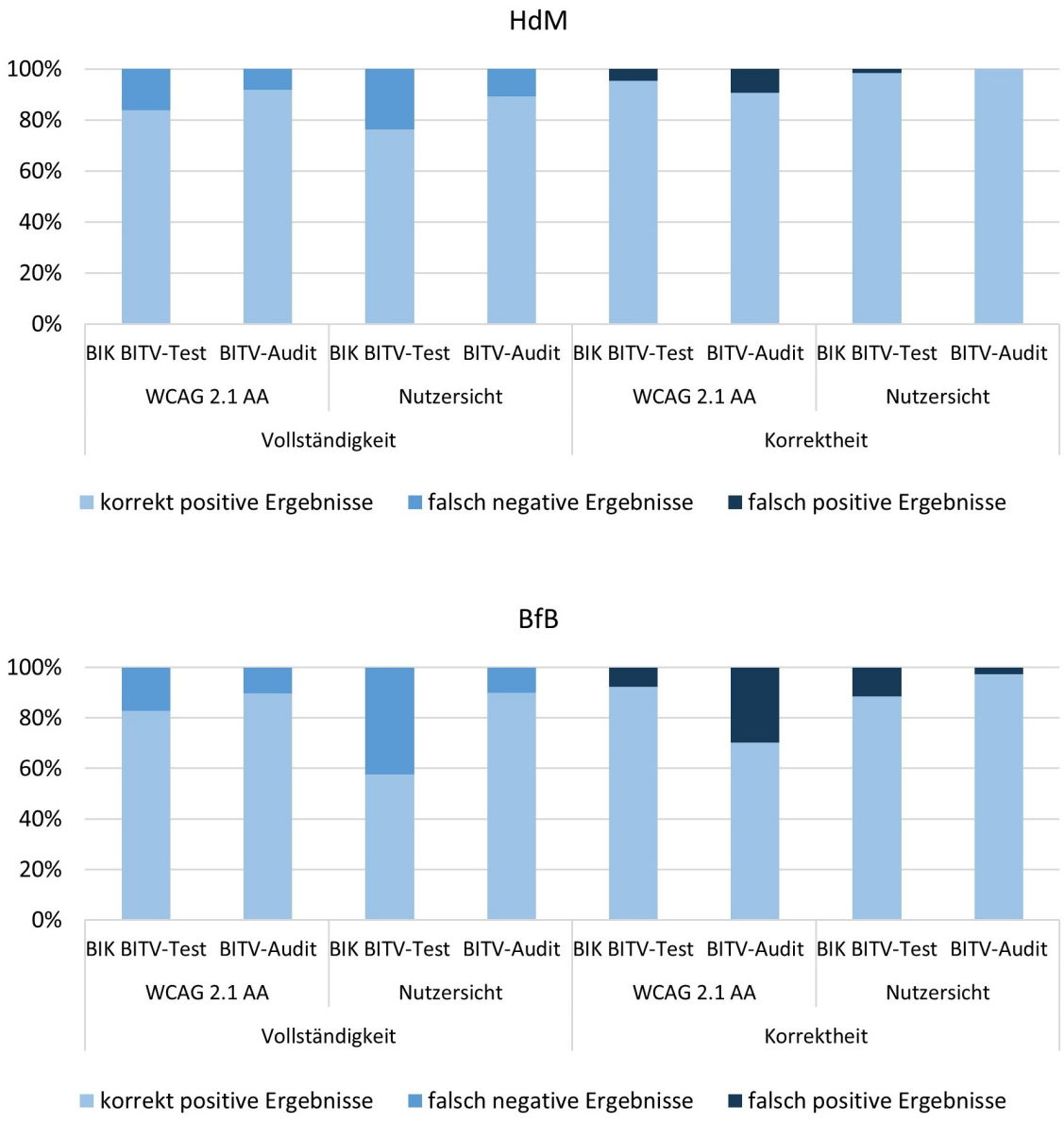


Abbildung 9.4: Vollständigkeit und Korrektheit der Prüfberichtergebnisse

Bei einer Differenz ($Diff_A$) von mindestens 0,10 zwischen den normalisierten Werten der beiden Websites HdM und BfB ($Diff_A \geq$ 0,10), werden Auffälligkeiten beim Prüfverfahren beobachtet. Diese Auffälligkeiten treten je nach Barrierefreiheit des Prüfgegenstands auf.

Der BIK BITV-Test hat keine Schwächen bei den betrachteten Metriken. Bei der folgenden Metrik liegen die Stärken des Prüfverfahrens: „Bewertung nach WCAG 2.1 AA“ ($MS = 0{,}83$). Der BIK BITV-Test zeigt bei der Metrik „Bewertung aus Nutzersicht“ Auffälligkeiten ($Diff_A = 0{,}18$). Die Vollständigkeit für die eher barrierefreie Website (BfB) ist schlechter als die Vollständigkeit der weniger barrierefreien Website (HdM).

Der BITV-Audit hat keine Schwächen bei den betrachteten Metriken. Bei den folgenden beiden Metriken liegen die Stärken des Prüfverfahrens: „Bewertung nach WCAG 2.1 AA“ ($MS = 0{,}91$) und „Bewertung aus Nutzersicht“ ($MS = 0,90$). Außerdem zeigt der BITV-Audit keine Auffälligkeiten.

Der Kriteriumscore für das Kriterium „Korrektheit“ beträgt für den BIK BITV-Test 0,94 und für den BITV-Audit 0,90. Die Kriteriumscores der „Korrektheit“ ($Diff = 0{,}04$; zugunsten BIK BITV-Test) unterscheiden sich weniger als die Kriteriumscores der „Vollständigkeit“ ($Diff = 0{,}15$; zugunsten BITV-Audit). Bei der Metrik „Bewertung aus Nutzersicht“ ($Diff = 0{,}06$) haben die Prüfverfahren große Gemeinsamkeiten. Bei keiner Metrik unterscheiden sich die Prüfverfahren stark.

Der BIK BITV-Test hat keine Schwächen bei den betrachteten Metriken. Bei den folgenden beiden Metriken liegen die Stärken des Prüfverfahrens: „Bewertung nach WCAG 2.1 AA“ ($MS = 0{,}94$) und „Bewertung aus Nutzersicht“ ($MS = 0{,}93$). Der BIK BITV-Test zeigt bei der Metrik „Bewertung aus Nutzersicht“ Auffälligkeiten ($Diff_A = 0{,}10$). Die Korrektheit der eher barrierefreien Website (BfB) ist schlechter als die Korrektheit der weniger barrierefreien Website (HdM).

Der BITV-Audit hat keine Schwächen bei den betrachteten Metriken. Bei den folgenden beiden Metriken liegen die Stärken des Prüfverfahrens: „Bewertung nach WCAG 2.1 AA“ ($MS = 0{,}80$) und „Bewertung aus Nutzersicht“ ($MS = 0{,}99$). Der BITV-Audit zeigt bei der Metrik „Bewertung nach WCAG 2.1 AA“ Auffälligkeiten ($Diff_A = 0{,}21$). Die Korrektheit der eher barrierefreien Website (BfB) ist schlechter als die Korrektheit der weniger barrierefreien Website (HdM).

Qualität des Prüfberichts

Tabelle 9.4 zeigt eine Übersicht der sechs Metriken, die zur Bewertung des Kriteriums „Qualität des Prüfberichts“ ($G = 6{,}67\%$) herangezogen werden. Die Bewertungen der Metriken „Verfügbarkeit von Screenshots“, „Verfügbarkeit von Markierungen in Screenshots“ und „Verfügbarkeit von dokumentierten Lösungen zu den Barrieren“ beruhen auf der in Unterabschnitt 9.1.2 beschriebenen Expertenwertung und den Prüfberichten. Es werden nur Probleme, die auch im Prüfbericht dokumentiert wurden und vom Prüfverfahren gewertet werden dürfen, berücksichtigt. Folglich werden Flüchtigkeitsfehler und fälschlicherweise bewertete Probleme nicht miteinbezogen. Hinweise ohne Auswirkungen auf die Bewertung werden ebenfalls nicht berücksichtigt. Die Bewertung der Metrik „Anpassbarkeit der Ausführlichkeit“ wird auf Basis der jeweiligen Verfahrensbeschreibungen (vgl. DIAS GmbH, 2019; T-Systems MMS, 2020b) sowie den erstellten Prüfberichten getroffen. Die „Verständlichkeit des Prüfberichts“ wird anhand von den bereits in Unterabschnitt 9.1.3 beschriebenen durchgeführten Interviews bewertet. Die „Barrierefreiheit des Prüfberichts“ beruht auf den PAC 3 Reports der Prüfberichte, diese sind im elektronischen Anhang zu finden. Alle Prüfberichte werden von PAC 3 als nicht PDF/UA-konform bewertet.

Bei den Interviews wurden folgende Gedanken von den Interviewpartnern zum BIK BITV-Test geäußert: Das Verständnis der Barrieren ist stark von den Formulierungen des Prüfers abhängig und ob eine Hilfestellung für die Lösung angeboten wird (I1, I4, I5, I6). Es gibt keine Möglichkeit Screenshots einzubinden, sodass die Beschreibung der Probleme ausführlicher ausfallen muss (I1, I4, I6). Grundsätzlich gibt es nur eingeschränkte Formatierungsoptionen (keine farbliche Hervorhebung) und die Einbindung von Quellcode ist sehr umständlich möglich (I5, I6). Für den Leser ist es hilfreich, dass der Prüfbericht nach Prüfschritten sortiert ist und diese nachgelesen werden können (I1).

Bei den Interviews wurden folgende Gedanken von den Interviewpartnern zum BITV-Audit geäußert: Der Prüfbericht bietet für jedes Problem eine Ursachenbeschreibung, eine Beschreibung der Auswirkung für Benutzergruppen, einen Screenshot und Empfehlungen zur Behebung (I2, I7). Die bisherige Erfahrung zeigt, dass die Probleme behoben werden und nur sehr selten Rückfragen zur Lösungsempfehlung gestellt werden (I2, I3).

Der Kriteriumscore für das Kriterium „Qualität des Prüfberichts“ beträgt für den BIK BITV-Test 0,30 und für den BITV-Audit 0,79. Bei der Metrik „Anpassbarkeit der Ausführlichkeit“ ($Diff = 0{,}00$) haben die Prüfverfahren große Gemeinsamkeiten. Bei folgenden drei Metriken unterscheiden sich die Prüfverfahren stark: „Verfügbarkeit von Screenshots“ ($Diff = 0{,}76$; zugunsten BITV-Audit), „Verfügbarkeit von Markierungen in Screenshots“

	BIK BITV-Test	**BITV-Audit**
Verfügbarkeit von Screenshots		
HdM	0,00	0,65
BfB	0,00	0,88
Metrikscore	**0,00**	**0,76**
Verfügbarkeit von Markierungen in Screenshots		
HdM	0,00	0,61
BfB	0,00	0,82
Metrikscore	**0,00**	**0,72**
Verfügbarkeit von dokumentierten Lösungen zu den Barrieren		
HdM	0,13	0,97
BfB	0,00	1,00
Metrikscore	**0,06**	**0,98**
Anpassbarkeit der Ausführlichkeit		
Ergebniszusammenfassung	1,00	1,00
Alternativversion	0,00	0,00
Metrikscore	**0,50**	**0,50**
Verständlichkeit des Prüfberichts		
Interview	0,54	0,89
Metrikscore	**0,54**	**0,89**
Barrierefreiheit des Prüfberichts		
HdM	0,72	0,89
BfB	0,72	0,89
Metrikscore	**0,72**	**0,89**
Kriteriumscore	**0,30**	**0,79**

Tabelle 9.4: Ergebnisse des Kriteriums „Qualität des Prüfberichts“

($Diff = 0{,}72$; zugunsten BITV-Audit) und „Verfügbarkeit von dokumentierten Lösungen zu den Barrieren“ ($Diff = 0{,}92$; zugunsten BITV-Audit).

Bei den folgenden drei Metriken liegen die Schwächen des BIK BITV-Test: „Verfügbarkeit von Screenshots“ ($MS = 0{,}00$), „Verfügbarkeit von Markierungen in Screenshots“ ($MS = 0{,}00$) und „Verfügbarkeit von dokumentierten Lösungen zu den Barrieren“ ($MS = 0{,}06$). Das Prüfverfahren hat keine Stärken bei den betrachteten Metriken. Der BIK BITV-Test zeigt bei der Metrik „Verfügbarkeit von dokumentierten Lösungen zu den Barrieren“ Auffälligkeiten ($Diff_A = 0{,}13$). Es werden bei der weniger barrierefreien Website (HdM) vereinzelt Empfehlungen zur Lösung gegeben, bei der eher barrierefreien Website (BfB) keine einzige.

Der BITV-Audit hat keine Schwächen bei den betrachteten Metriken. Bei den folgenden vier Metriken liegen die Stärken des Prüfverfahrens: „Verfügbarkeit von Screenshots“ ($MS = 0{,}76$), „Verfügbarkeit von dokumentierten Lösungen zu den Barrieren“ ($MS = 0{,}98$), „Verständlichkeit des Prüfberichts“ ($MS = 0{,}89$) und „Barrierefreiheit des Prüfberichts“ ($MS = 0{,}89$). Der BITV-Audit zeigt bei den Metriken „Verfügbarkeit von Screenshots“ ($Diff_A = 0,23$) und „Verfügbarkeit von Markierungen in Screenshots“ ($Diff_A = 0{,}21$) Auffälligkeiten. Der Anteil der Probleme mit Screenshots und der Anteil der Probleme mit Markierungen in Screenshots sind bei der eher barrierefreien Website (BfB) höher als bei der weniger barrierefreien Website (HdM).

Qualität der Stichprobe

Tabelle 9.5 zeigt eine Übersicht der beiden Metriken, die zur Bewertung des Kriteriums „Qualität der Stichprobe“ ($G = 5{,}96\%$) herangezogen werden. Die Bewertung der Metriken „Umfang der Stichprobe“ und „Varianz der Stichprobe“ werden auf Basis der jeweiligen Verfahrensbeschreibungen (DIAS GmbH, 2019; T-Systems MMS, 2020b) sowie dem Durchführungsbeschluss 2018/1524 (Europäische Kommission, 2018) getroffen. Bei der Bewertung des BITV-Audits wird zudem ein internes Dokument zur Bestimmung des Aufwands (T-Systems MMS, o.D.d) berücksichtigt.

Da es bei beiden Verfahrensbeschreibungen keine definierte maximale Webseitenanzahl gibt, ist dieser Aspekt bei beiden Prüfverfahren nicht beantwortbar und wird bei der Bewertung nicht berücksichtigt. Die minimale Webseitenanzahl für den BITV-Audit wurde in Absprache mit T-Systems MMS auf fünf Webseiten festgelegt: Startseite, Impressum, mindestens eine Webseite eines angebotenen Dienstes, EzB und eine Webseite mit einem anderen Erscheinungsbild, wenn vorhanden (A. Nebe [T-Systems MMS], persönliche

	BIK BITV-Test	BITV-Audit
Umfang der Stichprobe		
Minimale Webseitenanzahl	0,50	1,00
Maximale Webseitenanzahl	n.b.	n.b.
Variabilität anhand der Komplexität	1,00	1,00
Metrikscore	**0,75**	**1,00**
Varianz der Stichprobe		
Auswahl ist unabhängig vom Auftraggeber	1,00	1,00
Auswahl ist nicht vorhersehbar	1,00	1,00
Auswahl erfasst sämtliche Barrieren	1,00	1,00
Auswahl umfasst unterschiedliche Webseitentypen	1,00	1,00
Auswahl umfasst unterschiedliche Webseitenzustände	1,00	1,00
Auswahl deckt ganze Prozesse ab	1,00	1,00
Auswahl berücksichtigt Stellenwert einer Webseite für Benutzbarkeit des Webangebots	1,00	1,00
Auswahl umfasst unterschiedliche Funktionen	1,00	1,00
Auswahl nutzt das Zufallsprinzip	0,00	0,00
Auswahl umfasst Nicht-Web-Dokumente	0,00	1,00
Metrikscore	**0,71**	**0,86**
Kriteriumscore	**0,73**	**0,93**

Tabelle 9.5: Ergebnisse des Kriteriums „Qualität der Stichprobe“

Kommunikation, 02. März 2021). Bei der Metrik „Varianz der Stichprobe“ wird der Aspekt „Stellenwert für Benutzbarkeit des Webangebots“ für den BITV-Audit mit 1,00 bewertet, da dies durch den Anspruch auf Abdeckung der Haupt-Use-Cases gewährleistet wird.

Der Kriteriumscore für das Kriterium „Qualität der Stichprobe“ beträgt für den BIK BITV-Test 0,73 und für den BITV-Audit 0,93. Bei keiner Metrik haben die Prüfverfahren große Gemeinsamkeiten oder starke Unterschiede.

Der BIK BITV-Test hat keine Schwächen bei den betrachteten Metriken. Die Stärken des Prüfverfahrens liegen bei der Metrik „Umfang der Stichprobe“ ($MS = 0{,}75$).

Der BITV-Audit hat keine Schwächen bei den betrachteten Metriken. Bei den folgenden beiden Metriken liegen die Stärken des Prüfverfahrens: „Umfang der Stichprobe“ ($MS = 1{,}00$) und „Varianz der Stichprobe“ ($MS = 0{,}86$).

Qualitätssicherung

Tabelle 9.6 zeigt eine Übersicht der drei Metriken, die zur Bewertung des Kriteriums „Qualitätssicherung“ ($G = 5{,}58\%$) herangezogen werden. Die Bewertung der Metrik „Qualitätssicherung innerhalb eines Prüfauftrags“ wird auf Basis der jeweiligen Verfahrensbeschreibungen (vgl. DIAS GmbH, 2019; T-Systems MMS, 2020b) getroffen. Bei noch zu geringem Informationsgehalt einzelner Bewertungsaspekte wurde zusätzlich noch Rücksprache mit T-Systems MMS gehalten (A. Nebe [T-Systems MMS], persönliche Kommunikation, 02. März 2021). Die Bewertung der Metrik „Sicherstellung der Einheitlichkeit der Prüfung“ basiert zusätzlich zu den Verfahrensbeschreibungen (vgl. DIAS GmbH, 2019; T-Systems MMS, 2020b) auch auf der Inspektion des BITV-Test-Prüftools (DIAS GmbH, o.D.d) und einer persönlichen Kommunikation mit T-Systems MMS (A. Nebe [T-Systems MMS], persönliche Kommunikation, 01. März 2021). Zur Bewertung der Metrik „Sicherstellung der Qualifikation“ werden ebenfalls die Verfahrensbeschreibungen (vgl. DIAS GmbH, 2019; T-Systems MMS, 2020b) sowie ein Artikel über die Prüfstellen des BIK-Prüfverbunds (vgl. DIAS GmbH, o.D.c) betrachtet. Die Beurteilung des BIK BITV-Tests beruht zusätzlich auf einer persönlichen Kommunikation mit der DIAS GmbH (D. Fischer [DIAS GmbH], persönliche Kommunikation, 05. März 2021).

Der Aspekt „Anfängliche Begleitung durch einen erfahrenen Prüfer“ wird beim BIK BITV-Test mit 0,50 bewertet, da es keine direkte Begleitung durch einen Prüfer gibt. Durch die gemeinsame Abstimmung mit einem erfahrenen Prüfer besteht jedoch die Möglichkeit, Unsicherheiten abzustimmen und einen großen Lerneffekt zu erzielen. Auch beim BITV-Audit wurden noch unklare Aspekte in einer persönlichen Kommunikation mit T-Systems MMS abgestimmt (A. Nebe [T-Systems MMS], persönliche Kommunikation, 01. März 2021). Bei beiden Prüfverfahren werden keine formalen periodischen Prüfungen der Qualifikation durchgeführt. Beide erhalten jedoch die Wertung 0,50, da durch die kontinuierlichen Besprechungen der gefundenen Probleme sowie der Qualitätssicherung der Ergebnisse, alle Prüfer fortlaufend ihre Qualifikation untereinander abstimmen und verbessern.

Der Kriteriumscore für das Kriterium „Qualitätssicherung“ beträgt für den BIK BITV-Test 0,73 und für den BITV-Audit 0,60. Bei der Metrik „Sicherstellung der Einheitlichkeit der Prüfung“ ($Diff = 0{,}00$) haben die Prüfverfahren große Gemeinsamkeiten. Bei der Metrik „Qualitätssicherung innerhalb eines Prüfauftrags“ ($Diff = 0{,}60$; zugunsten BIK BITV-Test) unterscheiden sich die Prüfverfahren stark.

Der BIK BITV-Test hat keine Schwächen bei den betrachteten Metriken. Bei den folgenden beiden Metriken liegen die Stärken des Prüfverfahrens: „Qualitätssicherung innerhalb eines

	BIK BITV-Test	BITV-Audit
Qualitätssicherung innerhalb eines Prüfauftrags		
Webseitenauswahl durch mehrere Experten	1,00	0,00
Mehrere Prüfer führen Prüfung unabhängig durch	1,00	0,00
Teambesprechung der gefundenen Probleme	1,00	1,00
Qualitätssicherung durch qualifizierte Stelle	1,00	1,00
Gleicher Prüfer prüft kein identisches Webangebot	1,00	0,00
Metrikscore	**1,00**	**0,40**
Sicherstellung der Einheitlichkeit der Prüfung		
Zugriff auf veröffentlichte Prüfergebnisse	1,00	1,00
Zugriff auf alle Prüfergebnisse	0,00	1,00
Teambesprechung der gefundenen Probleme	1,00	1,00
statistische Auswertung der Einheitlichkeit	1,00	0,00
Metrikscore	**0,75**	**0,75**
Sicherstellung der Qualifikation eines Prüfers		
Qualifizierungsprozess um Prüfer zu werden	1,00	1,00
Anfängl. Begleitung durch erfahrenen Prüfer	0,50	1,00
Periodische Prüfung der Qualifikation	0,50	0,50
Regelmäßige Fortbildungsmaßnahmen	1,00	1,00
Festgelegte Anzahl von Tests pro Jahr	0,00	0,00
Statistische Auswertung der Qualifikation	0,00	0,00
Austausch mit Menschen mit Behinderung	0,00	1,00
Metrikscore	**0,43**	**0,64**
Kriteriumscore	**0,73**	**0,60**

Tabelle 9.6: Ergebnisse des Kriteriums „Qualitätssicherung“

	BIK BITV-Test	**BITV-Audit**
Qualifizierungsaufwand für Prüfer		
Interview	0,29	0,33
Metrikscore	**0,29**	**0,33**
Flüchtigkeitsfehler in Prüfberichten		
HdM	0,89	0,82
BfB	0,71	0,88
Metrikscore	**0,80**	**0,85**
Kriteriumscore	**0,54**	**0,59**

Tabelle 9.7: Ergebnisse des Kriteriums „Einfachheit der Prüfungsdurchführung“

Prüfauftrags“ ($MS = 1{,}00$) und „Sicherstellung der Einheitlichkeit der Prüfung“ ($MS = 0{,}75$).

Der BITV-Audit hat keine Schwächen bei den betrachteten Metriken. Bei der Metrik „Sicherstellung der Einheitlichkeit der Prüfung“ ($MS = 0{,}75$) liegen die Stärken des Prüfverfahrens.

Einfachheit der Prüfungsdurchführung

Tabelle 9.7 zeigt eine Übersicht der beiden Metriken, die zur Bewertung des Kriteriums „Einfachheit der Prüfungsdurchführung“ ($G = 2{,}44\%$) herangezogen werden. Die Metrik „Qualifizierungsaufwand für Prüfer“ wird anhand von den bereits in Unterabschnitt 9.1.3 beschriebenen durchgeführten Interviews bewertet. Die Metrik „Flüchtigkeitsfehler in Prüfberichten“ beruht auf der in Unterabschnitt 9.1.2 beschriebenen Expertenwertung.

Bei den Interviews wurden folgende Gedanken von den Interviewpartnern zum BIK BITV-Test geäußert: Das Kennenlernen der Standards, der Prüfschritte und des BITV-Test-Prüftools wird als sehr hoher Qualifizierungsaufwand eingeschätzt (I4, I5, I6). Der Qualifizierungsprozess (Durchführung von drei BITV-Tests mit einem erfahrenen Prüfer im Tandemverfahren) selbst wird von den Interviewpartnern als unterschiedlich aufwändig wahrgenommen (I1, I5, I6).

Bei den Interviews wurden folgende Gedanken von den Interviewpartnern zum BITV-Audit geäußert: Nach einer kurzen Einarbeitungszeit kann man mit einfachen Prüfungen beginnen und wird durch einen erfahrenen Prüfer unterstützt (I2). Das Durchdringen der

komplexen Thematik, aber auch das Kennenlernen und das Bedienen des Screenreaders benötigen dann jedoch viel mehr Zeit (I2, I3, I7).

Der Kriteriumscore für das Kriterium „Einfachheit der Prüfungsdurchführung“ beträgt für den BIK BITV-Test 0,54 und für den BITV-Audit 0,59. Bei beiden Metriken „Qualifizierungsaufwand für Prüfer“ ($Diff = 0{,}04$) und „Flüchtigkeitsfehler in Prüfberichten“ ($Diff = 0{,}05$) haben die Prüfverfahren große Gemeinsamkeiten.

Der BIK BITV-Test hat keine Schwächen bei den betrachteten Metriken. Bei der Metrik „Flüchtigkeitsfehler in Prüfberichten“ ($MS = 0{,}80$) liegen die Stärken des Prüfverfahrens. Der BIK BITV-Test zeigt bei der Metrik „Flüchtigkeitsfehler in Prüfberichten“ Auffälligkeiten ($Diff_A = 0{,}18$). Der Prüfbericht der eher barrierefreien Website (BfB) enthält einen höheren Anteil an Flüchtigkeitsfehler als der Prüfbericht der weniger barrierefreien Website (HdM).

Der BITV-Audit hat keine Schwächen bei den betrachteten Metriken und zeigt keine Auffälligkeiten. Bei der Metrik „Flüchtigkeitsfehler in Prüfberichten“ ($MS = 0{,}85$) liegen die Stärken des Prüfverfahrens. Der BITV-Audit zeigt keine Auffälligkeiten bei der Metrik „Flüchtigkeitsfehler in Prüfberichten“.

9.2.4 Effizienz der Prüfung

Die Ergebnisse folgender drei Kriterien werden in diesem Abschnitt zusammengefasst: „Aufwand“ ($G = 4{,}77\%$), „Skalierbarkeit hinsichtlich Prüfer“ ($G = 4,04\%$), „Grad der Toolunterstützung“ ($G = 3{,}87\%$). Die Kriteriumscores der aufgeführten Kriterien sind in Abbildung 9.5 zu sehen.

Bei dem Kriterium „Grad der Toolunterstützung“ ($Diff = 0{,}08$) haben die Prüfverfahren große Gemeinsamkeiten. Bei keinem Kriterium unterscheiden sich die Prüfverfahren stark.

Bei dem Kriterium „Aufwand“ ($KS = 0{,}17$) liegen die Schwächen des BIK BITV-Tests. Das Prüfverfahren hat keine Stärken bei den betrachteten Kriterien.

Der BITV-Audit hat keine Schwächen bei den betrachteten Kriterien. Bei dem Kriterium „Skalierbarkeit hinsichtlich Prüfer“ ($KS = 0{,}80$) liegen die Stärken des Prüfverfahrens. Die folgenden Abschnitte gehen genauer auf die einzelnen Kriterien ein.

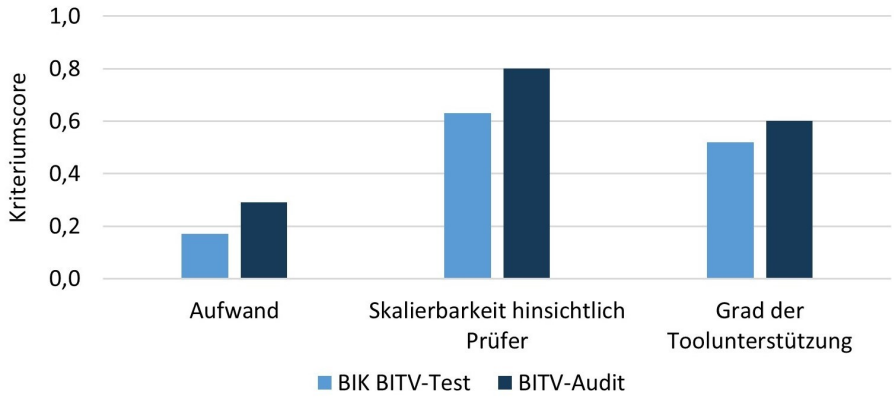


Abbildung 9.5: Ergebnisse der drei Kriterien, die die Effizienz betreffen

Aufwand

Tabelle 9.8 zeigt eine Übersicht der drei Metriken, die zur Bewertung des Kriteriums „Aufwand“ ($G = 4{,}77\%$) herangezogen werden. Die Metrik „Anzahl der Prüfschritte/kriterien“ beruht auf der jeweiligen Dokumentation der Prüfschritte/-kriterien. Beim BIK BITV-Test gibt es 60 Prüfschritte. Jeder dieser Prüfschritte ist ausführlich dokumentiert und gibt an was er genau prüft, wieso das wichtig ist und wie das Prüfvorgehen aussieht (vgl. DIAS GmbH, o.D.f). Beim BITV-Audit entspricht die Ebene des Prüfschrittes dem Prüfkriterium. Zu jedem Prüfkriterium gibt es Erläuterungen, die beschreiben was genau geprüft wird und wie bei der Prüfung des Kriteriums vorgegangen werden muss (vgl. T-Systems MMS, o.D.c). Das heißt, die verwendeten Begriffe „Prüfschritt” und „Prüfkriterium” der beiden Prüfverfahren haben große Gemeinsamkeiten und können hier in vergleichbarer Weise betrachtet werden. Der BITV-Audit hat insgesamt 85 Prüfkriterien. Damit werden alle WCAG 2.1 Erfolgskriterien abgedeckt, 31 der zusätzlichen EN 301 549 Anforderungen der Tab. A.1 sowie alle zusätzlichen Anforderungen der BITV 2.0. Im Prüfvorgehen des BITV-Audits ist die Auswahl der Prüfkriterien ein fester Prozessschritt. Da der BIK BITV-Test nur die WCAG 2.1 AA abdeckt, werden zur Auswertung der Metrik „Anzahl der Prüfschritte/-kriterien“ nur die 50 Prüfkriterien des BITV-Audits, die auf die WCAG 2.1 Erfolgskriterien der Konformitätsstufe A und AA abzielen, berücksichtigt. Die Bewertung der Metrik „Anzahl der involvierten Experten“ wird auf Basis der jeweiligen Verfahrensbeschreibungen (vgl. DIAS GmbH, 2019; T-Systems MMS, 2020b) getroffen. Die Bewertung der Metrik „Zeitaufwand für Prüfprozess“ basiert auf dem empirisch ermittelten Zeitaufwand bei der Erstellung der Prüfberichte. Eine detaillierte Aufschlüsselung der Zeiten ist im elektronischen Anhang zu finden.

	BIK BITV-Test	BITV-Audit
Anzahl der Prüfschritte/-kriterien		
Anzahl der Prüfschritte/-kriterien	60	50
Normalisierter Wert	0,00	0,17
Metrikscore	**0,00**	**0,17**
Anzahl der involvierten Experten		
Bei der Prüfung	2	1
Normalisierter Wert	0,00	1,00
Bei der Qualitätssicherung	1	2
Normalisierter Wert	1,00	0,00
Metrikscore	**0,50**	**0,50**
Zeitaufwand für Prüfprozess		
HdM - benötigte Zeit in Min.	835	688
Normalisierter Wert	0,00	0,18
BfB - benötigte Zeit in Min.	564	444
Normalisierter Wert	0,00	0,21
Metrikscore	**0,00**	**0,19**
Kriteriumscore	**0,17**	**0,29**

Tabelle 9.8: Ergebnisse des Kriteriums „Aufwand“

Der Kriteriumscore für das Kriterium „Aufwand“ beträgt für den BIK BITV-Test 0,17 und für den BITV-Audit 0,29. Bei der Metrik „Anzahl der involvierten Experten“ ($Diff$ = 0,00) haben die Prüfverfahren große Gemeinsamkeiten. Bei keiner Metrik unterscheiden sich die Prüfverfahren stark.

Bei den folgenden zwei Metriken liegen die Schwächen des BIK BITV-Tests: „Anzahl der Prüfschritte/-kriterien“ (MS = 0,00) und „Zeitaufwand für Prüfprozess“ (MS = 0,00). Das Prüfverfahren hat keine Stärken und keine Auffälligkeiten bei den betrachteten Metriken.

Bei den folgenden zwei Metriken liegen die Schwächen des BITV-Audits: „Anzahl der Prüfschritte/-kriterien“ (MS = 0,17) und „Zeitaufwand für Prüfprozess“ (MS = 0,19). Das Prüfverfahren hat keine Stärken und keine Auffälligkeiten bei den betrachteten Metriken.

Skalierbarkeit hinsichtlich Prüfer

Tabelle 9.9 zeigt eine Übersicht der drei Metriken, die zur Bewertung des Kriteriums „Skalierbarkeit hinsichtlich der Prüfer“ (G = 4,04%) herangezogen werden. Alle Metriken werden anhand von den bereits in Unterabschnitt 9.1.3 beschriebenen durchgeführten Interviews bewertet.

Bei den Interviews wurden folgende Gedanken von den Interviewpartnern zum BIK BITV-Test und dem Einsatz von Menschen mit Behinderung als Prüfer geäußert: Die Bewertung hängt stark von der Behinderung des Menschen ab (I1, I5). Je nach Einschränkung ist die Durchführung bestimmter Prüfschritte sehr schwierig oder auch gar nicht möglich (I1, I4, I5, I6). Die Erfahrungswerte, die ein Prüfer mit Behinderung in ein Team einbringen kann, werden jedoch als sehr wertvoll bewertet. Als Beispiel wurde der sichere Umgang mit einem Screenreader genannt (I1, I5, I6).

Bei den Interviews wurden folgende Gedanken von den Interviewpartnern zum BITV-Audit und dem Einsatz von Menschen mit Behinderung als Prüfer geäußert: Die Bewertung hängt stark von der Behinderung des Menschen ab (I2, I3, I7). Je nach Einschränkung ist die Prüfung bestimmter Prüfkriterien sehr schwierig oder auch gar nicht möglich (I2, I3, I7). Zudem wurde auch angemerkt, dass die für die Prüfung notwendigen Dokumente und Tools höchstwahrscheinlich nicht vollständig barrierefrei sind und dass dies auch ein Problem darstellen könnte (I2). Die Zusammenarbeit im Team mit anderen Prüfern,

	BIK BITV-Test	BITV-Audit
Einsatz von Menschen mit Behinderung		
Interview - einzelner Prüfer	0,13	0,39
Interview - Prüfer im Team	0,79	0,83
Metrikscore	**0,46**	**0,61**
Einsatz eines Teams von Prüfern mit unters. Rollen		
Interview	0,67	0,83
Metrikscore	**0,67**	**0,83**
Integration von Anfängern in ein Prüfteam		
Interview	0,75	0,95
Metrikscore	**0,75**	**0,95**
Kriteriumscore	**0,63**	**0,80**

Tabelle 9.9: Ergebnisse des Kriteriums „Skalierbarkeit hinsichtlich Prüfer“

wird als problemlos möglich eingeschätzt (I2, I7). Ferner werden die Erfahrungswerte von Menschen mit Behinderung als Bereicherung wahrgenommen (I3).

Bei den Interviews wurden folgende Gedanken von den Interviewpartnern zum BIK BITV-Test und dem Einsatz eines Teams mit unterschiedlichen Rollen geäußert: Standardmäßig gibt es schon zwei Rollen bei diesem Prüfverfahren: Prüfer und Qualitätssicherung. Dies wird vom Tool sehr gut unterstützt (I1, I5, I6). Eine Aufteilung der Prüfschritte nach verschiedenen Benutzergruppen ist im Tool derzeit nicht vorhanden, so müsste diese Aufteilung selbstständig vorgenommen und im Team außerhalb des Tools kommuniziert werden (I5, I6). Zudem gibt es auch den Fall, dass ein gefundenes Problem bereits bei einem vorherigen Prüfschritt gefunden wurde und dann nicht doppelt dokumentiert wird, sondern auf die vorherige Dokumentation verwiesen wird. Dies wäre dann nicht mehr so einfach möglich (I5).

Bei den Interviews wurden folgende Gedanken von den Interviewpartnern zum BITV-Audit und dem Einsatz eines Teams mit unterschiedlichen Rollen geäußert: Die Aufteilung muss exakt abgesprochen werden, ist aber bereits im Arbeitsalltag umgesetzt worden (I2, I3, I7). Das ist auch eine Möglichkeit, um neue Teammitglieder bei der Prüfung von komplexen Projekten zu integrieren (I2).

Bei den Interviews wurden folgende Gedanken von den Interviewpartnern zum BIK BITV-Test und der Integration von Anfängern in ein Prüfteam geäußert: Nach der Qualifizierung zählen auch wenig erfahrene Prüfer als vollwertiges Mitglied im BIK-Prüfverbund (I1, I5,

I6). Zudem bietet der Tandemtest hier eine Möglichkeit der Qualitätssicherung (I5). Es wurde aber auch erwähnt, dass wenig erfahrene Prüfer anfänglich einige Zeit aufwenden müssen, um sich mit den Hilfsmitteln und der Dokumentation richtig vertraut zu machen (I4, I6).

Bei den Interviews wurden folgende Gedanken von den Interviewpartnern zum BITV-Audit und der Integration von Anfängern in ein Prüfteam geäußert: Die Integration von Anfängern ist im Arbeitsalltag ein fester Bestandteil. So wird wenig erfahrenen Prüfern Zeit gegeben, um sich in die Thematik einzuarbeiten, aber auch schon nach kurzer Zeit an Projekten mitzuarbeiten (I2, I3). Durch ausführliche Dokumentationen aber auch den Austausch im Team können sie sich mit der Thematik vertraut machen und mit der Zeit das Prüfverfahren in Gänze verstehen (I7).

Der Kriteriumscore für das Kriterium „Skalierbarkeit hinsichtlich Prüfer“ beträgt für den BIK BITV-Test 0,63 und für den BITV-Audit 0,80. Bei keiner Metrik haben die Prüfverfahren große Gemeinsamkeiten oder Unterschiede. Bei der Metrik „Einsatz von Menschen mit Behinderung“ zeigen jeweils beide Prüfverfahren große Differenzen bei der Einschätzung als einzelner Prüfer und als Prüfer im Team. Die Einschätzung für einzelne Prüfer fällt viel geringer aus als die Einschätzung für Prüfer im Team.

Der BIK BITV-Test hat keine Schwächen bei den betrachteten Metriken. Bei der Metrik „Integration von Anfängern in ein Prüfteam“ ($MS = 0{,}75$) liegen die Stärken des Prüfverfahrens.

Der BITV-Audit hat keine Schwächen bei den betrachteten Metriken. Bei den folgenden zwei Metriken liegen die Stärken des Prüfverfahrens: „Einsatz eines Teams von Prüfern mit unterschiedlichen Rollen“ ($MS = 0{,}83$) und „Integration von Anfängern in ein Prüfteam“ ($MS = 0{,}95$).

Grad der Toolunterstützung

Tabelle 9.10 zeigt eine Übersicht der drei Metriken, die zur Bewertung des Kriteriums „Grad der Toolunterstützung“ ($G = 3{,}87\%$) herangezogen werden. Die Metrik „Multiuserfähigkeit“ wird auf Basis der Verfahrensbeschreibungen (vgl. DIAS GmbH, 2019; T-Systems MMS, 2020b) sowie den verwendeten Prüftools (vgl. DIAS GmbH, o.D.d; T-Systems MMS, o.D.c; T-Systems MMS, o.D.e) beantwortet. Außerdem wurde in der persönlichen Kommunikation mit T-Systems MMS abgesprochen, ob die Daten zentral gespeichert werden (A.Nebe [T-Systems MMS], persönliche Kommunikation, 01. März 2021).

	BIK BITV-Test	BITV-Audit
Multiuserfähigkeit		
Wechsel der Prüfer während der Prüfungsdurchführung ist vom Prüfverfahren angedacht	0,00	0,00
Wechsel der Prüfer während der Prüfungsdurchführung ist technisch einfach möglich	0,00	1,00
Zentrale Speicherung der Prüfinformationen	1,00	1,00
Metrikscore	**0,33**	**0,67**
Plattformunabhängigkeit		
Plattformunabhängigkeit	1,00	0,67
Metrikscore	**1,00**	**0,67**
Automatisierung von Abläufen		
Tool unterstützt Webseitenauswahl	0,00	0,00
Tool wird automatisch mit Informationen zur Barrierefreiheit des Prüfgegenstands befüllt	0,00	0,00
Tool unterstützt beim Editieren der Probleme	1,00	1,00
Tool passt Prüfschritte automatisch an	0,00	1,00
Interview	0,17	0,33
Metrikscore	**0,23**	**0,47**
Kriteriumscore	**0,52**	**0,60**

Tabelle 9.10: Ergebnisse des Kriteriums „Grad der Toolunterstützung“

Auch die Metriken „Plattformunabhängigkeit“ und „Automatisierung von Abläufen“ werden durch eine Inspektion der Prüftools (vgl. DIAS GmbH, o.D.d; T-Systems MMS, o.D.c; T-Systems MMS, o.D.e) bewertet. Bei der Metrik „Automatisierung von Abläufen“ werden zusätzlich die bereits in Unterabschnitt 9.1.3 beschriebenen durchgeführten Interviews berücksichtigt.

Bei den Interviews wurden folgende Gedanken von den Interviewpartnern zum BIK BITV-Test geäußert: Die Webseitenauswahl wird vom Prüfer manuell vorgenommen (I5). Die Prüfung selbst ist überhaupt nicht automatisiert. Es werden nur kleine Hilfsmittel eingesetzt, die den Prüfer bei der manuellen Prüfung unterstützen (I2, I4, I5, I6). Der Prüfer trägt die Inhalte auch selbst in das BITV-Test-Prüftool ein (I1, I4). Am Ende wird der Prüfbericht dann aus diesen Eingaben automatisch erstellt (I2, I4, I5, I6).

Bei den Interviews wurden folgende Gedanken von den Interviewpartnern zum BITV-Audit geäußert: Die Prüfung selbst ist überhaupt nicht automatisiert. Es werden nur kleine Hilfsmittel eingesetzt, die den Prüfer bei der manuellen Prüfung unterstützen (I3, I7). Das Tool hilft dem Prüfer bei der Erstellung des Prüfberichts. Die individuellen Problembeschreibungen müssen jedoch durch den Prüfer formuliert werden (I2, I3, I7). Eine Automatisierung der Prüfung wird derzeit erforscht, aber noch nicht eingesetzt (I2, I3).

Der Kriteriumscore für das Kriterium „Grad der Toolunterstützung“ beträgt für den BIK BITV-Test 0,52 und für den BITV-Audit 0,60. Bei keiner Metrik haben die Prüfverfahren große Gemeinsamkeiten oder Unterschiede.

Bei der Metrik „Automatisierung von Abläufen“ ($MS = 0{,}23$) liegen die Schwächen des BIK BITV-Tests. Bei der Metrik „Plattformunabhängigkeit“ ($MS = 1{,}00$) liegen die Stärken des Prüfverfahrens.

Der BITV-Audit hat weder Schwächen noch Stärken bei den betrachteten Metriken.

9.2.5 Rahmenbedingungen des Prüfverfahrens

Es werden die Ergebnisse folgender sieben Kriterien in diesem Abschnitt zusammengefasst: „Optionale Inputformate“ ($G = 5{,}57\%$), „Bandbreite an geprüften Benutzungskonstellationen“ ($G = 4{,}19\%$), „Öffentlichkeit des Prüfverfahrens“ ($G = 4{,}07\%$), „Lizenz“ ($G = 2{,}96\%$), „Zertifikat“ ($G = 2{,}90\%$), „Organisatorische Anforderungen“ ($G = 1{,}51\%$) und „Formate des Prüfberichts“ ($G = 1{,}48\%$). Die Kriteriumscores der aufgeführten Kriterien sind in Abbildung 9.6 zu sehen.

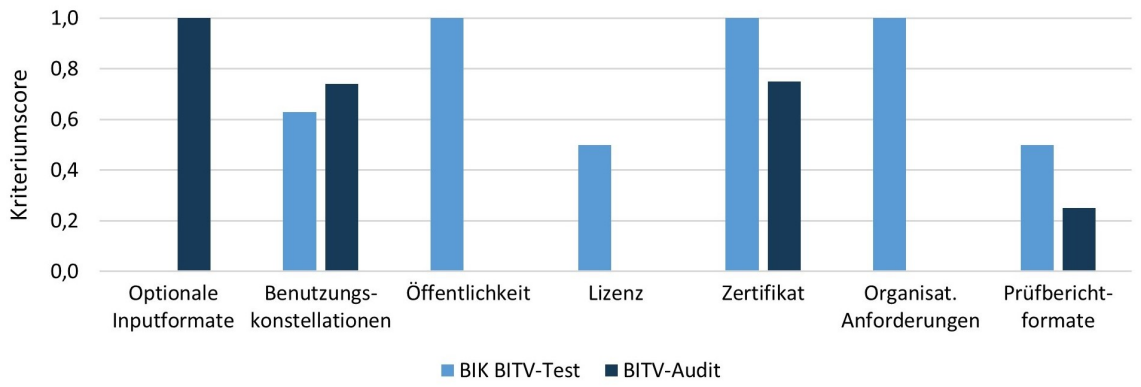


Abbildung 9.6: Ergebnisse der sieben Kriterien, die die Rahmenbedingungen des Prüfverfahrens betreffen

Bei keinem Kriterium haben die Prüfverfahren große Gemeinsamkeiten. Bei den folgenden vier Kriterien unterscheiden sich die Prüfverfahren stark: „Optionale Inputformate“ ($Diff = 1{,}00$; zugunsten BITV-Audit), „Öffentlichkeit des Prüfverfahrens“ ($Diff = 1{,}00$; zugunsten BIK BITV-Test), „Lizenz“ ($Diff = 0{,}50$; zugunsten BIK BITV-Test) sowie „Organisatorische Anforderungen“ ($Diff = 1{,}00$; zugunsten BIK BITV-Test).

Bei dem Kriterium „Optionale Inputformate“ ($KS = 0{,}00$) liegen die Schwächen des BIK BITV-Tests. Bei den folgenden drei Kriterien liegen die Stärken des Prüfverfahrens: „Öffentlichkeit des Prüfverfahrens“ ($KS = 1{,}00$), „Zertifikat“ ($KS = 1{,}00$) und „Organisatorische Anforderungen“ ($KS = 1{,}00$).

Bei den folgenden drei Kriterien liegen die Schwächen des BITV-Audits: „Öffentlichkeit des Prüfverfahrens“ ($KS = 0{,}00$), „Lizenz“ ($KS = 0{,}00$) und „Organisatorische Anforderungen“ ($KS = 0{,}00$). Bei den zwei Kriterien „Optionale Inputformate“ ($KS = 1{,}00$) und „Zertifikat“ ($KS = 0{,}75$) liegen die Stärken des Prüfverfahrens.

Optionale Inputformate

Tabelle 9.11 zeigt eine Übersicht der Metrik, die zur Bewertung des Kriteriums „Optionale Inputformate“ ($G = 5{,}57\%$) herangezogen wird. Die Metrik „Anzahl der optionalen Inputformate“ wird für den BIK BITV-Test anhand eines Artikels zu den Grenzen des Prüfverfahrens bewertet (vgl. DIAS GmbH, o.D.e). Für den BITV-Audit werden die Checkliste (T-Systems MMS, o.D.c) und ein Dokument zur Prüfung zeitbasierter Medien (T-Systems MMS, o.D.a) berücksichtigt.

	BIK BITV-Test	BITV-Audit
Anzahl der optionalen Inputformate		
PDF-Dokument	0,00	1,00
Textverarbeitungsdokument	0,00	1,00
Tabellenkalkulationsdokument	0,00	1,00
Präsentationsdatei	0,00	1,00
Videodatei	0,00	1,00
Audiodatei	0,00	1,00
Metrikscore	**0,00**	**1,00**
Kriteriumscore	**0,00**	**1,00**

Tabelle 9.11: Ergebnisse des Kriteriums „Optionale Inputformate

Der Kriteriumscore für das Kriterium „Optionale Inputformate“ beträgt für den BIK BITV-Test 0,00 und für den BITV-Audit 1,00. Bei diesem Kriterium unterscheiden sich die Prüfverfahren stark ($Diff = 1{,}00$; zugunsten BITV-Audit).

Bei diesem Kriterium liegen die Schwächen des BIK BITV-Tests ($MS = 0{,}00; KS = 0{,}00$). Bei diesem Kriterium liegen die Stärken des BITV-Audits ($MS = 1{,}00; KS = 1{,}00$).

Bandbreite an geprüften Benutzungskonstellationen

Tabelle 9.12 zeigt eine Übersicht der drei Metriken, die zur Bewertung des Kriteriums „Bandbreite an geprüften Benutzungskonstellationen“ ($G = 4{,}19\%$) herangezogen werden. Zur Bewertung der drei Metriken „Verwendete Plattformen“, „Verwendete Browser“ und „Verwendete Hilfsmittel“ werden die Werkzeugliste des BIK BITV-Tests (DIAS GmbH, o.D.h) sowie ein interner Demoprüfbericht des BITV-Audits (T-Systems MMS, 2019) genutzt und Rücksprache mit T-Systems MMS (A. Nebe [T-Systems MMS], persönliche Kommunikation, 02. März 2021) gehalten. Zusätzlich wird bei der Metrik „Verwendete Hilfsmittel“ eine Übersicht der Bookmarklets von T-Systems MMS (T-Systems MMS, o.D.b) berücksichtigt.

Beim BIK BITV-Test ist die Prüfumgebung genau definiert und immer identisch. Die Prüfumgebung beim BITV-Audit kann je nach Bedarf angepasst werden. Deswegen wird die Prüfumgebung in jedem Prüfbericht des BITV-Audits genau definiert. Diese unterschiedlichen Herangehensweisen spiegeln sich auch in den Bewertungen wieder. Beim BIK BITV-Test kann je nach Kundenwunsch auf Android oder iOS geprüft werden. Die beiden

	BIK BITV-Test	BITV-Audit
Verwendete Plattformen		
Windows	1,00	1,00
Mac	0,00	0,50
Linux	0,00	0,50
Android	1,00	0,50
iOS	1,00	0,50
Metrikscore	**0,60**	**0,60**
Verwendete Browser		
Chrome	1,00	1,00
Firefox	1,00	0,50
Safari	0,00	0,50
Edge	0,00	1,00
Internet Explorer	0,50	0,50
Opera	0,00	0,50
iOS Safari	1,00	0,50
Android Chrome	1,00	0,50
Samsung Internet	0,00	0,50
Metrikscore	**0,50**	**0,61**
Verwendete Hilfsmittel		
Screenreader	1,00	1,00
Bookmarklets	0,21	1,00
Onlinedienst	1,00	1,00
Browser- und BS-Funktionen	0,50	1,00
Browser-Erweiterungen	1,00	1,00
Sonstige Programme	1,00	1,00
Metrikscore	**0,78**	**1,00**
Kriteriumscore	**0,63**	**0,74**

Tabelle 9.12: Ergebnisse des Kriteriums „Bandbreite an geprüften Benutzungskonstellationen“

	BIK BITV-Test	BITV-Audit
Öffentlichkeit des Prüfverfahrens		
Öffentliche detaillierte Verfahrensbeschreibung	1,00	0,00
Öffentliche Diskussionsplattform für Prüfverfahren	1,00	0,00
Metrikscore	**1,00**	**0,00**
Kriteriumscore	**1,00**	**0,00**

Tabelle 9.13: Ergebnisse des Kriteriums „Öffentlichkeit des Prüfverfahrens“

Betriebssysteme haben den Wert 1,00 erhalten, da auf jeden Fall mit einer mobilen Ansicht geprüft wird. Beim BITV-Audit wird, wenn von Auftraggeber nicht anders gewünscht, auf Windows geprüft. Daher erhält dieses Betriebssystem den Wert 1,00. Es können aber auf Wunsch des Auftraggebers alle Betriebssysteme verwendet werden, deshalb erhalten diese den Wert 0,50.

Der BIK BITV-Test kann auf expliziten Wunsch des Auftraggebers auch mit dem Internet Explorer prüfen, daher erhält dieser Browser die Wertung 0,50. Beim BITV-Audit wird grundsätzlich mit Chrome und Edge geprüft, diese beiden Browser erhalten deswegen den Wert 1,00. Auf Kundenwunsch können jedoch auch alle anderen Browser verwendet werden, daher erhalten sie den Wert 0,50.

Bei der Bewertung des Aspekts „Browser- und Betriebssystem-Funktionen“ wird berücksichtigt, dass der BITV-Audit zusätzlich zur Zoomfunktion des Browsers auch die Zoomfunktion des Betriebssystems nutzt. Der BIK BITV-Test verwendet ausschließlich die Zoomfunktion des Browsers (vgl. DIAS GmbH, 2017).

Bei beiden Prüfverfahren werden nicht alle ausgewählten Browser zur Prüfung aller Prüfschritte eingesetzt. Außerdem werden auch nicht alle ausgewählten Browser mit allen ausgewählten Hilfsmitteln kombiniert eingesetzt.

Der Kriteriumscore für das Kriterium „Bandbreite an geprüften Benutzungskonstellationen“ beträgt für den BIK BITV-Test 0,63 und für den BITV-Audit 0,74. Bei der Metrik „Verwendete Plattformen“ ($Diff = 0{,}00$) haben die Prüfverfahren große Gemeinsamkeiten. Bei keiner Metrik unterscheiden sich die Prüfverfahren stark.

Der BIK BITV-Test hat keine Schwächen bei den betrachteten Metriken. Bei der Metrik „Verwendete Hilfsmittel“ ($MS = 0{,}78$) liegen die Stärken des Prüfverfahrens.

	BIK BITV-Test	BITV-Audit
Lizenzbedingungen		
Lizenzbedingungen	0,50	0,00
Metrikscore	**0,50**	**0,00**
Kriteriumscore	**0,50**	**0,00**

Tabelle 9.14: Ergebnisse des Kriteriums „Lizenz“

Der BITV-Audit hat keine Schwächen bei den betrachteten Metriken. Bei der Metrik „Verwendete Hilfsmittel“ ($MS = 1{,}00$) liegen die Stärken des Prüfverfahrens.

Öffentlichkeit des Prüfverfahrens

Tabelle 9.13 zeigt eine Übersicht der Metrik, die zur Bewertung des Kriteriums „Öffentlichkeit des Prüfverfahrens“ ($G = 4{,}07\%$) herangezogen wird. Zum BIK BITV-Test gibt es eine detaillierte öffentliche Verfahrensbeschreibung (DIAS GmbH, 2019) und auch eine ausführliche Dokumentation der Prüfschritte (DIAS GmbH, o.D.f). Zudem wird die Weiterentwicklung des Verfahrens öffentlich auf Github diskutiert (vgl. DIAS GmbH, o.D.a). Durch Rücksprache mit T-Systems MMS hat sich herausgestellt, dass es keine öffentliche, detaillierte Verfahrensbeschreibung sowie Dokumentation der Prüfkriterien gibt und auch keine öffentliche Diskussionsplattform (A. Nebe [T-Systems MMS], persönliche Kommunikation, 13. April 2021).

Der Kriteriumscore für das Kriterium „Öffentlichkeit des Prüfverfahrens“ beträgt für den BIK BITV-Test 1,00 und für den BITV-Audit 0,00. Bei diesem Kriterium unterscheiden sich die Prüfverfahren stark ($Diff = 1{,}00$; zugunsten des BIK BITV-Test).

Bei diesem Kriterium liegen die Stärken des BIK BITV-Tests ($MS = 1{,}00; KS = 1{,}00$). Bei diesem Kriterium liegen die Schwächen des BITV-Audits ($MS = 0{,}00; KS = 0{,}00$).

Lizenz

Tabelle 9.14 zeigt eine Übersicht der Metrik, die zur Bewertung des Kriteriums „Lizenz“ ($G = 2{,}96\%$) herangezogen wird. Der BIK BITV-Test kann als Selbstbewertung für interne Zwecke von jedem angewendet werden. Ein abschließender Test (inkl. Zertifizierung) kann von offiziellen Prüfstellen des BIK-Prüfverbunds durchgeführt werden. Um sich als

	BIK BITV-Test	BITV-Audit
Vorhandensein eines Prüfzeichens		
Prüfzeichen in Prüfbericht enthalten	1,00	1,00
Prüfzeichen kann auf Website eingebunden werden	1,00	1,00
Metrikscore	**1,00**	**1,00**
Veröffentlichung der Prüfergebnisse		
Veröffentlichung auf zentraler Website	1,00	0,00
Veröffentlichung auf eigener Website	1,00	1,00
Metrikscore	**1,00**	**0,50**
Kriteriumscore	**1,00**	**0,75**

Tabelle 9.15: Ergebnisse des Kriteriums „Zertifikat“

Organisation als Prüfstelle zu qualifizieren, gibt es bestimmte Auflagen und einen Qualifizierungsprozess für die Prüfer der Prüfstelle (vgl. DIAS GmbH, o.D.b; DIAS GmbH, o.D.c). Daher wird der BIK BITV-Test mit 0,50 bewertet. Beim BITV-Audit gibt es keinen vergleichbaren offenen Prüfverbund. Es gibt zwar Partner in Unterauftragsverhältnissen, T-Systems MMS bleibt jedoch der Halter des Prüfverfahrens (A. Nebe [T-Systems MMS], persönliche Kommunikation, 01. März 2021). Daher erhält das Prüfverfahren den Wert 0,00.

Der Kriteriumscore für das Kriterium „Lizenz“ beträgt für den BIK BITV-Test 0,50 und für den BITV-Audit 0,00. Bei diesem Kriterium unterscheiden sich die Prüfverfahren stark ($Diff = 0{,}50$; zugunsten des BIK BITV-Test).

Der BIK BITV-Test hat weder Stärken noch Schwächen bei diesem Kriterium. Bei diesem Kriterium liegen die Schwächen des BITV-Audits ($MS = 0{,}00; KS = 0{,}00$).

Zertifikat

Tabelle 9.15 zeigt eine Übersicht der beiden Metriken, die zur Bewertung des Kriteriums „Zertifikat“ ($G = 2,90\%$) herangezogen werden. Zur Bewertung der Metriken „Vorhandensein eines Prüfzeichens“ und „Veröffentlichung der Prüfergebnisse“ werden die Verfahrensbeschreibungen (DIAS GmbH, 2019; T-Systems MMS, 2020b) und Prüfberichte herangezogen. Zusätzlich werden beim BIK BITV-Test auch zwei Artikel zur Verwendung der Prüfzeichen berücksichtigt: Das Prüfzeichen des BIK BITV-Tests darf auf der Website des Auftraggebers veröffentlicht werden, muss aber auf den Prüfbericht oder die EzB verlinken

(vgl. DIAS GmbH, o.D.g; DIAS GmbH, 2020c). Durch eine Rücksprache mit T-Systems MMS stellte sich heraus, dass das Prüfzeichen des BITV-Audits auch vom Auftraggeber auf der eigenen Website eingebunden werden kann. Aber eine Veröffentlichung der Prüfergebnisse auf einer zentralen Website der T-Systems MMS ist nicht möglich (A. Nebe [T-Systems MMS], persönliche Kommunikation, 01. März 2021). Der Auftraggeber kann zusätzlich zum Prüfbericht eine Bescheinigung zur Konformität erhalten (vgl. T-Systems MMS, 2021b).

Der Kriteriumscore für das Kriterium „Zertifikat“ beträgt für den BIK BITV-Test 1,00 und für den BITV-Audit 0,75. Bei der Metrik „Vorhandensein eines Prüfzeichens“ haben die Prüfverfahren große Gemeinsamkeiten ($Diff = 0{,}00$). Bei der Metrik „Veröffentlichung der Prüfergebnisse“ ($Diff = 0{,}50$; zugunsten BIK BITV-Test) unterscheiden sich die Prüfverfahren stark.

Der BIK BITV-Test hat keine Schwächen bei den betrachteten Metriken. Bei den folgenden beiden Metriken liegen die Stärken des Prüfverfahrens: „Vorhandensein eines Prüfzeichens“ ($MS = 1{,}00$) und „Veröffentlichung der Prüfergebnisse“ ($MS = 1{,}00$).

Der BITV-Audit hat keine Schwächen bei den betrachteten Metriken. Bei der Metrik „Vorhandensein eines Prüfzeichens“ ($MS = 1{,}00$) liegen die Stärken des Prüfverfahrens.

Organisatorische Anforderungen

Tabelle 9.16 zeigt eine Übersicht der beiden Metriken, die zur Bewertung des Kriteriums „Organisatorische Anforderungen“ ($G = 1{,}51\%$) herangezogen werden. Zur Bewertung des finanziellen und zeitlichen Aufwands wurde Rücksprache mit der DIAS GmbH gehalten (D. Fischer [DIAS GmbH], persönliche Kommunikation, 05.03.2021). Für die Qualifizierung eines Prüfers fallen einmalig Kosten an. Für die Nutzung des BITV-Test-Prüftools fallen Betriebskosten an, welche einem prozentualen Anteil des Testpreises entsprechen. Der Testpreis variiert je nach Umfang der Webseitenauswahl und der Komplexität der ausgewählten Webseiten. Die Abrechnung erfolgt einmal pro Quartal. Zudem gibt es eine Grundgebühr pro Jahr. Diese Grundgebühr verringert sich jedoch um die in diesem Jahr gezahlten Betriebskosten. Wenn beispielsweise ein BIK BITV-Test in einem Jahr durchgeführt wurde, verringert sich die Grundgebühr um den Betriebskostenanteil dieses einen Tests. Es gibt keinen festen Zeitrahmen für den Qualifizierungsprozess. Die Dauer ist abhängig von den Fähigkeiten des Prüfers und der Verfügbarkeit von Prüfaufträgen. Je nach Auftragslage kann es zwischen wenigen Monaten oder auch einem halben bis ganzen Jahr dauern. Im Gegensatz zum BIK BITV-Test kann eine Organisation sich nicht als

	BIK BITV-Test	BITV-Audit
Finanzieller Aufwand		
Qualifizierungskosten	1,00	0,00
Grundbetrag	1,00	0,00
Betriebskosten	1,00	0,00
Metrikscore	**1,00**	**0,00**
Zeitlicher Aufwand		
Zeitraum	1,00	0,00
Metrikscore	**1,00**	**0,00**
Kriteriumscore	**1,00**	**0,00**

Tabelle 9.16: Ergebnisse des Kriteriums „Organisatorische Anforderungen“

Prüfstelle für den BITV-Audit qualifizieren. Es gibt keinen vergleichbaren Prüfverbund (A. Nebe [T-Systems MMS], persönliche Kommunikation, 16. März 2021). Daher erhält der BITV-Audit bei beiden Metriken den Wert 0,00 und der BIK BITV-Test den Wert 1,00.

Der Kriteriumscore für das Kriterium „Organisatorische Anforderungen“ beträgt für den BIK BITV-Test 1,00 und für den BITV-Audit 0,00. Bei beiden Metriken „Finanzieller Aufwand zur Qualifikation einer Organisation“ ($Diff = 1{,}00$; zugunsten BIK BITV-Test) und „Zeitlicher Aufwand zur Qualifikation einer Organisation“ ($Diff = 1{,}00$; zugunsten BIK BITV-Test) unterscheiden sich die Prüfverfahren stark.

Der BIK BITV-Test hat Stärken bei folgenden beiden Metriken: „Finanzieller Aufwand“ ($MS = 1{,}00$) und „Zeitlicher Aufwand“ ($MS = 1{,}00$).

Der BITV-Audit hat Schwächen bei beiden Metriken: „Finanzieller Aufwand“ ($MS = 0{,}00$) und „Zeitlicher Aufwand“ ($MS = 0{,}00$).

Formate des Prüfberichts

Tabelle 9.17 zeigt eine Übersicht der Metrik, die zur Bewertung des Kriteriums „Formate des Prüfberichts“ ($G = 1{,}48\%$) herangezogen wird. Die Bewertung der Metrik „Anzahl der Prüfberichtformate“ wird auf Basis der jeweiligen Verfahrensbeschreibungen (DIAS GmbH, 2019; T-Systems MMS, 2020b) getroffen.

	BIK BITV-Test	BITV-Audit
Anzahl der Prüfberichtformate		
PDF-Dokument	1,00	1,00
Textverarbeitungsdokument	0,00	0,00
Tabellenkalkulationsdokument	0,00	0,00
HTML-Webseite	1,00	0,00
Metrikscore	**0,50**	**0,25**
Kriteriumscore	**0,50**	**0,25**

Tabelle 9.17: Ergebnisse des Kriteriums „Formate des Prüfberichts“

Der Kriteriumscore für das Kriterium „Formate des Prüfberichts“ beträgt für den BIK BITV-Test 0,50 und für den BITV-Audit 0,25. Die Prüfverfahren haben weder große Gemeinsamkeiten noch starke Unterschiede. Auch hat keines der Prüfverfahren Stärken oder Schwächen bei diesem Kriterium.

9.3 Diskussion

9.3.1 In den Vergleich eingeflossene Expertise

Insgesamt waren neun Experten aus dem Bereich Barrierefreiheit an der Identifikation der Kriterien beteiligt, acht dieser Experten trugen auch zur Gewichtung der Kriterien bei. Ein qualifizierter Prüfer untersuchte die Startseiten zweier Websites mit beiden Prüfverfahren auf Barrierefreiheit, so dass vier Prüfberichte für den Vergleich bereitstanden. In den Prüfprozess des BITV-Audits war noch eine weitere Prüferin der T-Systems MMS involviert, beim BIK BITV-Tests waren es zwei weitere Prüfer des BIK-Prüfverbunds. Zudem haben sechs Barrierefreiheitsexperten an den durchgeführten Interviews teilgenommen und vier Barrierefreiheitsexperten nahmen am Workshop zur Klassifikation der im Prüfbericht dokumentierten Probleme teil.

9.3.2 Kriterienkatalog

Im Kapitel der verwandten Arbeiten wurde bereits dargelegt, dass ein umfassender Vergleich anhand realer Daten, die durch Experten erstellt werden, derzeit nicht existiert.

Brajnik (2008b) hat einzelne Kriterien zur Untersuchung von Prüfmethoden vorgeschlagen und in einer Studie verwendet: Korrektheit, Vollständigkeit, Reliabilität, Effizienz und Nützlichkeit. Im Rahmen dieser Arbeit wurde ein umfassender, generischer Kriterienkatalog erstellt, der auf der Expertise diverser Barrierefreiheitsexperten basiert. Der Kriterienkatalog besteht aus 22 Kriterien mit insgesamt 41 Metriken und ist somit umfassender als existierende Kriterienkataloge in verwandten Arbeiten.

Folgende zehn Kriterien haben die höchsten Gewichte bei der Expertengewichtung erhalten: Die „Abdeckung von WCAG 2.1 AA“, die „Abdeckung von zusätzlichen Kriterien aus EN 301 549 Tab. A.1“ und „Abdeckung von zusätzlichen Kriterien aus BITV 2.0“ haben hohe Gewichte erhalten. Die Anforderungen dieser Standards sind in Deutschland für den öffentlichen Sektor verpflichtend festgelegt. Da der zugrunde gelegte Standard für die Ergebnisse einer Konformitätsprüfung maßgeblich ist, könnten diese Kriterien als besonders wichtig erachtet werden. Da das Prüfen von Nicht-Web-Dokumenten in den Standards enthalten ist, könnte dies die hohe Gewichtung der „optionalen Inputformate“ erklären. Die hohen Werte für die „Vollständigkeit“, die „Korrektheit“ und die „Qualitätssicherung“ überraschen nicht, da es hier um die Aussagekraft der Ergebnisse geht. Die valide Aussage über die Barrierefreiheit der geprüften Webseiten stellt das Ziel einer manuellen Prüfung dar. Die hohe Gewichtung der „Qualität der Stichprobe“ lässt sich dadurch erklären, dass die Konformitätsaussage der geprüften Webseiten repräsentativ für die ganze Website sein soll. Eine Prüfung der ganzen Website ist nicht machbar, sodass die Auswahl der Webseiten sehr wichtig ist. Auch der „Aufwand“ wird von den Experten als wichtiges Kriterium wahrgenommen. Da eine Prüfung, die ideale Ergebnisse liefert, aber in der Realität nicht umsetzbar ist, keine gute Herangehensweise darstellt. Des Weiteren wird mit einer manuellen Prüfung das Ziel verfolgt die Barrierefreiheit einer Website nicht nur zu bestimmen, sondern auch zur Verbesserung der Barrierefreiheit beizutragen. Daher könnten die Experten das Kriterium „Qualität des Prüfberichts“ für wichtig empfunden haben.

Schließlich wurde der Vergleich zweier Prüfverfahren anhand dieses Kriterienkatalogs auf realen Daten durchgeführt. Die Ergebnisse des Vergleichs zeigen, dass verschiedene Dimensionen der Prüfverfahren mit dem Kriterienkatalog erfasst und quantifiziert werden können.

9.3.3 Gemeinsamkeiten und Unterschiede der Prüfverfahren

Tabelle 9.18 zeigt eine Übersicht der Gemeinsamkeiten und Unterschiede des BIK BITV-Tests und BITV-Audit. Anhand der Ergebnisse können folgende Gemeinsamkeiten identi-

Gemeinsamkeiten	- Abdeckung der WCAG 2.1 AA - Korrektheit - Grad der Toolunterstützung - Einfachheit der Prüfungsdurchführung
Unterschiede	- Abdeckung von zusätzl. Kriterien aus EN 301 549 Tab. A.1 - Abdeckung von zusätzl. Kriterien aus BITV 2.0 - Abdeckung von zusätzl. Kriterien aus WCAG 2.1 AAA - Optionale Inputformate - Öffentlichkeit des Prüfverfahrens - Lizenz - Organisatorische Anforderungen

Tabelle 9.18: Gemeinsamkeiten und Unterschiede der Prüfverfahren

fiziert werden: Beide Prüfverfahren decken alle WCAG 2.1 Erfolgskriterien der Konformitätsstufe A und AA ab. Außerdem finden sich nur sehr geringfügige Abweichungen bei der Korrektheit, dem Grad der Toolunterstützung sowie der Einfachheit der Prüfungsdurchführung.

Bei folgenden Aspekten liegen große Unterschiede zwischen den Prüfverfahren: Der BITV-Audit deckt eine Vielzahl an weiteren Anforderungen ab. Darunter zählen 31 der 42 zusätzlichen Kriterien der EN 301 549 Tabelle A.1, die zusätzlichen Anforderungen der BITV 2.0 und die WCAG 2.1 Erfolgskriterien der Konformitätsstufe AAA. Der BIK BITV-Test deckt diese zusätzlichen Kriterien nicht ab.

Zudem können alle gängigen optionalen Inputformate mit dem BITV-Audit geprüft werden. Der BIK BITV-Test prüft ausschließlich Websites auf Barrierefreiheit. Nicht-Web-Dokumente (z.B. PDF-Dokumente) sind von der Prüfung ausgeschlossen.

Der BIK BITV-Test ist vollständig offengelegt. Die einzelnen Prüfschritte sind im Detail beschrieben und öffentlich zugänglich. Das Prüfverfahren des BITV-Audits hingegen ist nicht öffentlich. Die Detailbeschreibungen der Prüfschritte sind nur für die T-Systems MMS zugänglich.

Während der BIK BITV-Test von diversen Prüfstellen des BIK-Prüfverbunds durchgeführt werden darf, wird der BITV-Audit ausschließlich von der T-Systems MMS durchgeführt. Dadurch kann sich auch keine Organisation qualifizieren um das Prüfverfahren anwenden

	Schwächen	**Stärken**
BIK BITV-Test	- Abdeckung von zusätzl. Kriterien aus EN 301 549 Tab. A.1 - Abdeckung von zusätzl. Kriterien aus BITV 2.0 - Abdeckung von zusätzl. Kriterien aus WCAG 2.1 AAA - Optionale Inputformate - Aufwand	- Abdeckung der WCAG 2.1 AA - Vollständigkeit - Korrektheit - Öffentlichkeit des Prüfverfahrens - Zertifikat - Organisatorische Anforderungen
BITV-Audit	- Öffentlichkeit des Prüfverfahrens - Lizenz - Organisatorische Anforderungen	- Abdeckung der WCAG 2.1 AA - Abdeckung von zusätzl. Kriterien aus BITV 2.0 - Abdeckung von zusätzl. Kriterien aus WCAG 2.1 AAA - Vollständigkeit - Korrektheit - Qualität des Prüfberichts - Qualität der Stichprobe - Optionale Inputformate - Skalierbarkeit hinsichtlich Prüfer - Zertifikat

Tabelle 9.19: Schwächen und Stärken der Prüfverfahren

zu können. Für den BIK BITV-Test kann sich eine Organisation unter bestimmten Auflagen qualifizieren und wird anschließend als offizielle BIK BITV-Test Prüfstelle in den BIK-Prüfverbund aufgenommen.

9.3.4 Schwächen und Stärken der Prüfverfahren

Anhand der Ergebnisse des quantitativen Vergleichs können für jedes Prüfverfahren Schwächen und Stärken identifiziert werden. Diese sind in Tabelle 9.19 zusammengefasst. Beim BIK BITV-Test werden die folgenden Schwächen identifiziert: Die fehlende Abdeckung der zusätzlichen Kriterien aus EN 301 549, BITV 2.0 und WCAG 2.1 AAA. Die fehlende Möglichkeit auch optionale Inputformate wie PDF-Dokumente zu prüfen, stellt eine wei-

tere Schwäche dar. Zuletzt ist noch der erhöhte Aufwand hinsichtlich der Prüfschritte und benötigten Zeiten zu nennen.

Es zeigen sich aber auch folgende Stärken des Prüfverfahrens: Die Abdeckung der WCAG 2.1 Erfolgskriterien mit der Konformitätsstufe A und AA ist vollständig. Das Prüfverfahren erreicht sehr hohe Werte im Bereich der Vollständigkeit und Korrektheit. Das heißt, sehr viele der vorhandenen Probleme werden vom Prüfverfahren als solche erkannt und dokumentiert. Es stellen auch nur sehr wenige dokumentierte Probleme eigentlich gar kein Problem dar. Des Weiteren ist hervorzuheben, dass das Prüfverfahren in Gänze veröffentlicht ist und öffentlich diskutiert werden kann. Außerdem schneidet das Prüfverfahren in den Bereichen Zertifikat und organisatorische Anforderungen sehr gut ab.

Bei den Kriterien Vollständigkeit, Korrektheit, Qualität des Prüfberichts und Einfachheit des Prüfvorgehens werden Auffälligkeiten bei mindestens einer Metrik beobachtet. Die weniger barrierefreie Website (HdM) erzielt bei allen Kriterien bessere Ergebnisse im Vergleich zur barrierefreieren Website (BfB).

Beim BITV-Audit werden die folgenden Schwächen identifiziert: Das Prüfverfahren kann von der Öffentlichkeit nicht vollumfänglich eingesehen werden und daher auch nicht öffentlich diskutiert werden. Da das Prüfverfahren nur von der T-Systems MMS durchgeführt werden darf, schneidet es im Bereich Lizenz sehr schlecht ab. Auch im Bereich der organisatorischen Anforderungen wird es sehr schlecht bewertet, da es nicht möglich ist, sich als Organisation für das Prüfverfahren zu qualifizieren.

Es zeigen sich aber auch folgende Stärken des Prüfverfahrens: Die sehr gute Abdeckung der zusätzlichen Kriterien der EN 301 549, BITV 2.0 und WCAG 2.1 AAA. Das Prüfverfahren erreicht sehr hohe Werte im Bereich der Vollständigkeit und Korrektheit. Zudem sticht das Prüfverfahren bei der Qualität des Prüfberichts mit umfangreichen Informationen wie Screenshots und Lösungsempfehlungen positiv heraus. Des Weiteren besteht die Möglichkeit alle gängigen optionalen Inputformate prüfen zu können. In Experteninterviews stellte sich zudem heraus, dass die Skalierbarkeit hinsichtlich der Prüfer bei diesem Prüfverfahren sehr gut ausfällt. Auch in den Bereichen Qualität der Stichprobe und Zertifikat schneidet das Prüfverfahren sehr gut ab.

Bei den Kriterien Korrektheit und Qualität des Prüfberichts werden Auffälligkeiten bei mindestens einer Metrik beobachtet. Die weniger barrierefreie Website (HdM) erzielt bei der Korrektheit bessere Ergebnisse im Vergleich zur barrierefreieren Website (BfB). Bei der Qualität des Prüfberichts schneidet die barrierefreiere Website (BfB) besser ab als die weniger barrierefreie Website (HdM).

9.3.5 Fazit des Vergleichs

Der BITV-Audit schneidet unter der durch die Experten ermittelten Gewichtung besser ab als der BIK BITV-Test. Eine individuelle Gewichtung kann je nach Situation und Zielsetzung notwendig sein und von der Expertengewichtung abweichen. In diesem Fall kann die Gewichtung problemlos an die eigenen Bedürfnisse adaptiert werden und ein neuer Gesamtscore für jedes Prüfverfahren ermittelt werden.

9.3.6 Probleme nach WCAG 2.1 und aus Nutzersicht

Der zugrunde liegende Standard ist maßgeblich für die Qualität der Ergebnisse einer Konformitätsprüfung. Diese Standards haben Schwächen und können nachweislich nicht alle für Benutzer bestehenden Probleme identifizieren (vgl. Calvo et al., 2016; Freire, 2012). Diese Problematik zeigte sich auch in den im Rahmen dieser Arbeit erstellten Expertenwertungen. Insgesamt wurden 23 gefundene Probleme nicht durch die WCAG 2.1 AA als solche erkannt, obwohl sie für Benutzer reale Probleme darstellen. Das heißt, dass tatsächliche Benutzerprobleme durch die WCAG 2.1 nicht vollständig abgedeckt werden.

9.3.7 Wichtige Implikationen der Ergebnisse

Es zeigte sich, dass die konformitätsbasierte Prüfung sicherstellt, dass eine diverse Basis an Problemen vollständig und korrekt erkannt werden kann. Dabei ist eine möglichst große Abdeckung der kritischen Standards durch das Prüfverfahren wichtig. In den Ergebnissen des Vergleichs wird jedoch auch deutlich, dass mit einem empirischen Prüfvorgehen die Vollständigkeit noch weiter gesteigert werden kann. Dadurch können Probleme erkannt werden, die nicht durch die Richtlinien abgedeckt werden. Die Kombination von empirischer Evaluation mit nachgelagerter Konformitätsprüfung, wie es der BITV-Audit einsetzt, verspricht hier gute Ergebnisse zur Erfassung der Barrierefreiheit einer Website. Die Kombination von Konformitätsprüfung und empirischer Evaluation soll auch bei der Gestaltung der WCAG 3.0 verfolgt werden (vgl. W3C, 2021).

Zudem sind ausführliche Prüfberichte mit Screenshots und Lösungsempfehlungen sehr wichtig, um ein umfassendes Verständnis beim Leser für Barrieren zu schaffen und so die Behebung der Barrieren zu erleichtern. Hier sollten unbedingt benutzerfreundliche Tools entwickelt werden, die die Prüfer bei der Erstellung von Prüfberichten effizient unterstützen.

Zudem zeigte sich in den Interviews, dass der Qualifizierungsaufwand, um ein erfahrener Prüfer zu werden, nicht zu unterschätzen ist und die Prüfungsdurchführung nicht trivial ist. Gute Fortbildungsprogramme und ein gemeinsamer Austausch können hier unterstützend wirken.

Der BIK BITV-Test mit seinem Tandemverfahren hat bei der Qualitätssicherung besser abgeschnitten als der BITV-Audit. Dieses Vorgehen scheint aber nicht zu bedeutend besseren Ergebnissen bzgl. Korrektheit und Vollständigkeit zu führen. Daher könnte das Tandemverfahren bei der Prüfung abgeschafft werden und dafür eine umfassende Qualitätssicherung der Ergebnisse dieses Fehlen kompensieren. Dies würde den Aufwand des Prüfverfahrens reduzieren. Das wurde auch bereits beim BIK BITV-Test mit der neuen Version umgesetzt (vgl. DIAS GmbH, 2020b).

Ein zentraler und wichtiger Aspekt im Bereich der manuellen Prüfmethoden ist die Skalierbarkeit der Prüfverfahren. Websites werden immer komplexer und die zu prüfende Webseitenauswahl ist ausschlaggebend für die Aussagekraft der Barrierefreiheit für die ganze Website. Der Umfang ist aber auch kritisch für die Durchführbarkeit solcher manuellen Prüfungen. Mithilfe folgender Idee könnte der Umfang zu einer bewältigbaren Größe reduziert werden und trotzdem eine valide Aussage über die Barrierefreiheit der ganzen Website getroffen werden: Die Auswahl der Stichprobe umfasst nicht ganze Webseiten, sondern die für die Barrierefreiheit kritischen Teilbereiche der einzelnen Webseiten. Der Arbeitskreis der WCAG 3.0 beschäftigt sich auch mit dieser Problematik. Wie sich in den Interviews mit Experten im Rahmen dieser Arbeit zeigt, kann die Zusammenarbeit in Teams hilfreich sein, um Prüfungen komplexer Webangebote effizient durchzuführen. So könnten beispielsweise Experten für verschiedene Benutzergruppen bei einer Prüfung zusammenarbeiten, aber auch Experten für verschiedene Teilbereiche wie die Prüfung von PDF-Dokumenten. Allem voran muss aber, wo möglich, der Grad der Toolunterstützung sowie die Automatisierung von Abläufen und Prüfschritten erhöht werden. Die Prüfer brauchen Tools, die sie bei der Prüfungsdurchführung und Dokumentation effizient unterstützen und entlasten.

9.3.8 Limitationen

Die vorliegende Arbeit wurde unter den gegebenen zeitlichen und finanziellen Umständen bestmöglich umgesetzt. Zur Einordnung der Forschungsergebnisse sollten folgende Limitationen berücksichtigt werden: Die Identifikation und Gewichtung der 22 Kriterien

wurde durch Barrierefreiheitsexperten vorgenommen, die Metriken der einzelnen Kriterien haben jedoch keine expertenbasierte Gewichtung, sondern tragen alle gleichwertig zum Kriteriumscore bei. Eine expertenbasierte Gewichtung dieser Metriken könnte zu einem abweichenden Ergebnis führen. Es wurde nur jeweils eine Webseite der Websites geprüft. Für eine Prüfung, die zu einer offiziellen Zertifizierung führen soll, muss eine repräsentative Stichprobe der Website durch den Prüfer ausgewählt werden. Dies war aufgrund des damit einhergehenden höheren Aufwands im Rahmen dieser Masterarbeit nicht möglich. Außerdem wurden alle Prüfungen nur durch einen Prüfer durchgeführt und es wurden nur zwei Websites geprüft. Es wäre noch besser gewesen, wenn eine Vielzahl an Prüfern jeweils mehrere unterschiedliche Websites geprüft hätte, um so eine noch allgemeingültigere Aussage zu erhalten. Beide Websites wurden mit der Zielsetzung ausgesucht, möglichst diverse Webinhalte einzuschließen, sodass alle Prüfschritte der Prüfverfahren anwendbar sind. Allerdings konnten nicht alle Webinhalte abgebildet werden, was in weiteren Arbeiten berücksichtigt werden könnte. Beispielsweise waren keine Audiodateien auf den geprüften Webseiten vorhanden. Des Weiteren handelt es sich bei beiden Websites um informationsorientierte Websites. Die Ergebnisse könnten für Prüfungen von komplexen Webanwendungen abweichen.

10 Zusammenfassung

In den letzten Jahren konnte eine zunehmende Digitalisierung in allen Lebensbereichen beobachtet werden. Zudem ist die Anzahl der Menschen mit Behinderung in unserer Gesellschaft kontinuierlich angestiegen. Daher hat die Relevanz der Barrierefreiheit von Informationstechnik im Laufe der letzten Jahren stark zugenommen. Mittels der Richtlinie 2016/2102 wurde die Umsetzung der Barrierefreiheit im öffentlichen Sektor für alle EU-Mitgliedsstaaten bereits verpflichtend festgelegt. Automatisierte Prüfverfahren können zwar einen ersten Eindruck über die Barrierefreiheit von Websites liefern, eine umfassende Bewertung der Barrierefreiheit ist derzeit aber nur durch manuelle Prüfmethoden möglich. Die zu Beginn betrachteten verwandten Arbeiten untersuchen die Prüfmethoden mehr auf einer konzeptionellen Ebene oder stellen eigene Methoden vor. Ein umfassender systematischer Vergleich bestehender manueller Prüfmethoden anhand von realen Daten, die durch Barrierefreiheitsexperten erstellt wurden, fehlt hingegen. In dieser Arbeit wurde deshalb hierfür eine Methodik entwickelt.

Die im Rahmen dieser Arbeit untersuchte Forschungsfrage lautet: „Welche Unterschiede lassen sich bei konformitätsbasierten und empirisch ermittelten manuellen Prüfmethoden zur Bewertung der Barrierefreiheit von Websites feststellen?“ Um einen systematischen quantitativen Vergleich im Rahmen dieser Arbeit durchführen zu können, wurde ein umfangreicher generischer Kriterienkatalog erstellt, der auf dem Fachwissen diverser Barrierefreiheitsexperten basiert. Der Kriterienkatalog kann unabhängig vom jeweiligen Prüfverfahren zur Analyse von manuellen Prüfmethoden verwendet werden. Mithilfe dieses erstellten Kriterienkatalogs wurden folgende zwei Prüfverfahren hinsichtlich ihrer Eignung und Wirksamkeit verglichen: Der BIK BITV-Test als eines der bekanntesten Konformitätsprüfungsverfahren in Deutschland sowie der BITV-Audit der T-Systems MMS als Vertreter der empirisch ermittelten Prüfmethoden.

Im Vergleich schließt der BITV-Audit anhand des definierten Kriterienkatalogs besser ab als der BIK BITV-Test. Jedoch ist zu beachten, dass es keine für alle Situationen allgemeingültige Gewichtung geben kann, die allen individuellen Anforderungen gerecht wird. Daher kann der Leser bei Bedarf eine individuelle Gewichtung festlegen und anhand dieser

das Ergebnis des Vergleichs für beide Verfahren neu berechnen. So kann die vorliegende Arbeit den Leser bei der Entscheidungsfindung unterstützen, welches Prüfverfahren in der jeweiligen persönlichen Situation geeigneter ist.

Außerdem zeigen die Ergebnisse folgende große Gemeinsamkeiten der beiden Prüfverfahren: Sie decken die WCAG 2.1 Erfolgskriterien der Konformitätsstufe A und AA vollständig ab. Zudem weisen sie ähnliche Werte im Bereich der Toolunterstützung und Komplexität der Prüfungsdurchführung auf. Starke Unterschiede stellten sich in folgenden Bereichen heraus: in der Abdeckung zusätzlicher Kriterien aus diversen Standards, im Umfang der optionalen Inputformate, in der Öffentlichkeit des Prüfverfahrens, in den Lizenzbedingungen und den organisatorischen Anforderungen, um das Prüfverfahren anwenden zu dürfen. Eine detaillierte Übersicht der Schwächen und Stärken der beiden Prüfverfahren ist in Unterabschnitt 9.3.4 zu finden. Außerdem hat sich auch im Verlauf dieser Arbeit gezeigt, dass die WCAG 2.1 nicht alle Probleme als solche erkennen, die jedoch nach Experteneinschätzungen für reale Benutzer bestehen.

Aus den Ergebnissen können interessante Erkenntnisse gewonnen werden. Eine der zukunftsweisenden ist die Kombination aus empirischem Prüfvorgehen und nachgelagerter Konformitätsprüfung. Dieses Vorgehen verspricht eine umfassende Prüfung der Barrierefreiheit, die über die Grenzen der Standards hinausgeht und den Mensch in den Mittelpunkt stellt. Die Skalierbarkeit der manuellen Prüfmethoden ist ein weiterer wichtiger Aspekt. Sie ist von mehreren Faktoren abhängig: der Auswahl der zu prüfenden Webseiten, der Prüfung als Teamarbeit mit Spezialisten und der Toolunterstützung sowie der Automatisierung von Abläufen und Prüfschritten.

Die vorliegende Arbeit erweitert den aktuellen Stand der Wissenschaft im Bereich der manuellen Prüfmethoden deutlich.

11 Ausblick

Im Folgenden werden wichtige Aspekte für künftige Arbeiten dargestellt, die im Rahmen dieser Arbeit deutlich wurden. Zukünftige Arbeiten könnten den Vergleich der Prüfverfahren für komplexe Webanwendungen anstelle von informationsorientierten Websites durchführen, um so zu überprüfen, ob sich die Ergebnisse unterscheiden. Daraus resultierend könnte eine Empfehlung des Prüfverfahrens abhängig von der Art des Prüfgegenstands abgeleitet werden.

In dieser Arbeit wurde zusammen mit Experten sorgfältig ein Kriterienkatalog basierend auf 22 Kriterien und gängigen Standards definiert. Dennoch ist es denkbar, dass in Zukunft weitere Kriterien und Metriken von hoher Relevanz sein könnten. Die Ergebnisse dieser Arbeit können dabei stets als Grundlage für mögliche künftige Erweiterungen in diesem Bereich dienen. Im Rahmen dieser Arbeit wurde vornehmlich die Perspektive der Prüfer und Organisationen, die die Prüfungen durchführen, eingenommen. In einem zukünftigen Forschungsprojekt könnte die Sichtweise des Auftraggebers stärker einbezogen und die Kriterien erweitert werden, um so ein umfassendes Bild zu vermitteln.

Wie bereits schon in den vorherigen Kapiteln angemerkt, ist es nicht möglich, eine allgemeingültige Gewichtung der Kriterien festzulegen, die für jede individuelle Situation exakt zugeschnitten ist. Es wäre jedoch spannend zu untersuchen, ob es verschiedene Interessensgruppen gibt, für die jeweils eine repräsentative Gewichtung festgelegt werden kann. Beispiele für solche Gruppen könnten die Folgenden sein: Ein Auftraggeber möchte eine Zertifizierung seiner Website vornehmen, da die Letzte bereits ein Jahr her ist und er eine aktuelle für die EzB braucht. Er weiß jedoch, dass in einem halben Jahr die Website komplett überarbeitet wird und dann erneut eine Zertifizierung ansteht. In diesem Fall könnte die Vollständigkeit und Korrektheit nicht ganz so relevant für den Auftraggeber sein wie der für die Prüfung notwendige Aufwand. Die Gewichtung könnte sich deutlich unterscheiden von einem Auftraggeber, der einen kompletten Relaunch der Website durchgeführt hat und nun die Barrierefreiheit dieser vollkommen neuen Version bewertet haben möchte. Hier liegt die Vermutung nahe, dass die Vollständigkeit und Korrektheit diesem Auftraggeber wichtiger sein wird als der dafür notwendige Aufwand.

Langfristig gesehen müssen weitere Forschungsarbeiten durchgeführt werden, um zu untersuchen, wie die manuelle Prüfung durch Tools effizient unterstützt werden kann und der benötigte Aufwand dadurch reduziert werden kann. Ein wichtiger Aspekt ist hierbei die Prüfungsdurchführung. Aber auch die Erstellung der Prüfberichte muss durch benutzerfreundliche Anwendungen erleichtert werden. Für die Automatisierung der manuellen Prüfung wird zum Beispiel die Forschung im Bereich der KI entscheidend sein. Beispielsweise stellt nicht die einfache Prüfung auf Vorhandensein eines Alternativtexts die kritische und aufwändige Herausforderung dar, sondern vielmehr die Entscheidung, ob ein Alternativtext in diesem Kontext notwendig ist und wie die Qualität dieses Alternativtexts zu bewerten ist. Erst wenn diese und ähnliche Entscheidungen verlässlich von einem automatischen Prüftool getroffen werden können, können die manuellen Prüfungen durch den Einsatz dieser Automatisierung wirklich beschleunigt werden.

Zusammenfassend kann festgehalten werden, dass es im Bereich der Bewertung der Barrierefreiheit von Websites noch ein großes Spektrum an spannenden Forschungsfragen gibt. Diese zu beantworten wird vor dem Hintergrund der weiter ansteigenden Anzahl der Menschen mit Behinderung sowie der voranschreitenden Digitalisierung essentiell sein.

Danksagung

An dieser Stelle möchte ich gerne meinen Dank aussprechen, da das Schreiben dieser Masterarbeit ohne die Unterstützung folgender Personen in dieser Form nicht möglich gewesen wäre.

Mein großer Dank gilt Herrn Prof. Dr. Gottfried Zimmermann und Herrn Tobias Ableitner für ihre zuverlässige Unterstützung während des gesamten Betreuungszeitraums, ihre konstruktive Kritik und ihre richtungweisenden Anregungen.

Gerne möchte ich mich bei Frau Anne-Marie Nebe von der Firma T-Systems Multimedia Solutions GmbH für die gute Kooperation bedanken. Die freundliche und verlässliche Zusammenarbeit hat mir wertvolle Einblicke gewährt und mir eine praxisnahe Forschung ermöglicht.

Mein Dank gilt auch den Expertinnen und Experten, die durch ihr Engagement, ihre Auskunftsbereitschaft und ihre interessanten Beiträge meine Masterarbeit wesentlich mitgeprägt haben.

Danken möchte ich meinem guten Freund Dr. Manuel Fritz für das Korrekturlesen meiner Masterarbeit und die motivierenden Worte.

Letztlich richte ich auch ein Dankeschön an meine Eltern, da sie mir das Studium ermöglicht haben und stets ein offenes Ohr für mich hatten.

Stuttgart, den 06. Mai 2021

A Literaturverzeichnis

1&1 IONOS SE. (2021). *Was ist eine Web-App? Definition und Web-App-Beispiele.* https://www.ionos.de/digitalguide/websites/web-entwicklung/verschiedene-app-formate-was-ist-eine-web-app/. (zuletzt abgerufen am 19.04.2021)

Abascal, J., Arrue, M. & Valencia, X. (2019). Tools for web accessibility evaluation. In Y. Yesilada & S. Harper (Hrsg.), *Web accessibility* (2. Aufl., S. 479 - 503). London: Springer.

Abduganiev, S. G. (2017). Towards automated web accessibility evaluation: A comparative study. *International Journal of Information Technology and Computer Science*, *9*, 18-44.

Aktion Mensch und Stiftung Digitale Chancen. (o.D.). *Flugpause.* https://www.biene-award.de/. (zuletzt abgerufen am 19.04.2021)

Baazeem, I. S. & Al-Khalifa, H. S. (2015). Advancements in web accessibility evaluation methods: How far are we? In *Proceedings of the 17th international conference on information integration and web-based applications & services.* New York, NY, USA: Association for Computing Machinery. Zugriff auf https://doi.org/10.1145/2837185.2843850 doi: 10.1145/2837185.2843850

Baumann, K. (2012). *Barrierefreiheit von Facebook - Untersuchung mit Hilfe des BITV-Tests* (Diplomarbeit). Hochschule Mittweida.

BIK für alle. (o.D.). *Automatisierte Tests.* https://bik-fuer-alle.de/automatisierte-tests.html. (zuletzt abgerufen am 24.02.2021)

Brajnik, G. (2005). Accessibility assessments through heuristic walkthroughs. *SIGCHI-Italy*, 77.

Brajnik, G. (2006). Web accessibility testing: When the method is the culprit. In *International conference on computers for handicapped persons* (S. 156–163).

Brajnik, G. (2008a, 09). Beyond conformance: The role of accessibility evaluation methods. In (S. 63-80). doi: 10.1007/978-3-540-85200-1_9

Brajnik, G. (2008b). A comparative test of web accessibility evaluation methods. In *Proceedings of the 10th international acm sigaccess conference on computers and*

accessibility (S. 113–120). New York, NY, USA: Association for Computing Machinery. Zugriff auf https://doi.org/10.1145/1414471.1414494 doi: 10.1145/1414471.1414494

Brajnik, G. (2009). Validity and reliability of web accessibility guidelines. In *Proceedings of the 11th international acm sigaccess conference on computers and accessibility* (S. 131–138). New York, NY, USA: Association for Computing Machinery. Zugriff auf https://doi.org/10.1145/1639642.1639666 doi: 10.1145/1639642.1639666

Brajnik, G. & Lomuscio, R. (2007). Samba: A semi-automatic method for measuring barriers of accessibility. In *Proceedings of the 9th international acm sigaccess conference on computers and accessibility* (S. 43–50). New York, NY, USA: Association for Computing Machinery. Zugriff auf https://doi.org/10.1145/1296843.1296853 doi: 10.1145/1296843.1296853

Brajnik, G., Mulas, A. & Pitton, C. (2007). Effects of sampling methods on web accessibility evaluations. In *Proceedings of the 9th international acm sigaccess conference on computers and accessibility* (S. 59–66). New York, NY, USA: Association for Computing Machinery. Zugriff auf https://doi.org/10.1145/1296843.1296855 doi: 10.1145/1296843.1296855

Bundesministerium für Arbeit und Soziales. (2019a). *Gemeinsam einfach machen.* https://www.einfach-teilhaben.de/DE/AS/Home/_functions/Weiterfuehrende_Infos_Dokumente/Artikel_1_GEM/artikel1.html. (zuletzt abgerufen am 21.04.2021)

Bundesministerium für Arbeit und Soziales. (2019b). *Verordnung zur Schaffung barrierefreier Informationstechnik nach dem Behindertengleichstellungsgesetz (Barrierefreie-Informationstechnik-Verordnung - BITV 2.0).*

Bundesministerium für Arbeit und Soziales. (2019c). *Zahlen und Fakten zur Lebenssituation von Menschen mit Behinderungen.* https://www.gemeinsam-einfach-machen.de/SharedDocs/Downloads/DE/AS/UN_BRK/Zahlen_und_Fakten.html. (zuletzt abgerufen am 05.01.2021)

Burkard, A., Zimmermann, G. & Schwarzer, B. (2021). Monitoring systems for checking websites on accessibility. *Frontiers in Computer Science*, *3*, 2. Zugriff auf https://www.frontiersin.org/article/10.3389/fcomp.2021.628770 doi: 10.3389/fcomp.2021.628770

Calvo, R., Seyedarabi, F. & Savva, A. (2016). Beyond web content accessibility guidelines: Expert accessibility reviews. In *Proceedings of the 7th international conference on software development and technologies for enhancing accessibility and fighting info-exclusion* (S. 77–84). New York, NY, USA: Association for Computing

Machinery. Zugriff auf https://doi.org/10.1145/3019943.3019955 doi: 10.1145/3019943.3019955

Colwell, C. & Petrie, H. (2001, Juli). Evaluation of guidelines for designing accessible web content. *SIGCAPH Comput. Phys. Handicap.* (70), 11–13. Zugriff auf https://doi.org/10.1145/501078.501082 doi: 10.1145/501078.501082

DIAS GmbH. (2017). *Prüfschritt 1.4.4a - Text auf 200% vergrößerbar.* https://testen.bitv-test.de/index.php?a=di&iid=87&s=n. (zuletzt abgerufen am 04.03.2021)

DIAS GmbH. (2019). *Beschreibung des Prüfverfahrens.* https://www.bitvtest.de/bitv_test/das_testverfahren_im_detail/verfahren.html. (zuletzt abgerufen am 14.10.2020)

DIAS GmbH. (2020a). *Abschließenden BITV-Test beauftragen.* https://www.bitvtest.de/bitv_test/bitv_test_beauftragen/infos_preise/preise_abschliessend.html. (zuletzt abgerufen am 14.10.2020)

DIAS GmbH. (2020b). *Anpassung des BITV-/WCAG-Test Prüfverfahrens 2021.* https://www.bitvtest.de/bitv_test/bitv_test_beauftragen/infos_preise/preise_abschliessend.html. (zuletzt abgerufen am 04.03.2021)

DIAS GmbH. (2020c). *Anpassung des BITV-WCAG-Test Prüfverfahrens 2021.* https://www.bitvtest.de/bitv_test/das_testverfahren_im_detail/vertiefend/infothek/artikel0/lesen/anpassung_2021.html. (zuletzt abgerufen am 04.03.2021)

DIAS GmbH. (2020d). *BITV-Test: Kurzvorstellung.* https://www.bitvtest.de/bitv_Test/einfuehrung/kurzvorstellung. (zuletzt abgerufen am 14.10.2020)

DIAS GmbH. (2020e). *Informationen und Preise.* https://www.bitvtest.de/bitv_test/bitv_test_beauftragen/infos_preise.html. (zuletzt abgerufen am 14.10.2020)

DIAS GmbH. (2021). *Überarbeitung von Prüfschritten im Jahr 2021.* https://bitvtest.de/bitv_test/das_testverfahren_im_detail/vertiefend/ueberarbeitung/2021.html. (zuletzt abgerufen am 05.04.2021)

DIAS GmbH. (o.D.a). *Bik test development c/o dias.* https://github.com/BIK-BITV. (zuletzt abgerufen am 16.04.2021)

DIAS GmbH. (o.D.b). *BITV-Selbstbewertungs-Formular.* https://www.bitvtest.de/bitv_test/bitv_test_selbst_anwenden/selbstbewertung.html. (zuletzt abgerufen am 14.10.2020)

DIAS GmbH. (o.D.c). *BITV-Test Prüfstellen.* https://www.bitvtest.de/bitv

_test/bitv_test_beauftragen/pruefstellen.html. (zuletzt abgerufen am 14.10.2020)

DIAS GmbH. (o.D.d). *BITV-Test-Prüftool.* https://testen.bitv-test.de/index.php?a=uo. (Internes Prüftool eingesehen am 08.01.2021)

DIAS GmbH. (o.D.e). *Die BITV 2.0 – Was prüft der BITV-Test, was prüft er nicht?* https://www.bitvtest.de/bitv_test/das_testverfahren_im_detail/vertiefend/die_bitv_20_was_prueft_der_bitv_test_was_prueft_er_nicht.html. (zuletzt abgerufen am 14.10.2020)

DIAS GmbH. (o.D.f). *Liste der Prüfschritte.* https://testen.bitv-test.de/index.php?a=dl. (zuletzt abgerufen am 16.03.2021)

DIAS GmbH. (o.D.g). *Neue BIK-Prüfzeichen für Webangebote.* https://www.bitvtest.de/bitv_test/das_testverfahren_im_detail/vertiefend/bik_pruefzeichen.html. (zuletzt abgerufen am 14.10.2020)

DIAS GmbH. (o.D.h). *Werkzeugliste.* https://www.bitvtest.de/bitv_test/das_testverfahren_im_detail/werkzeugliste.html. (zuletzt abgerufen am 14.10.2020)

Europäische Kommission. (2018). *Durchführungsbeschluss (EU) 2018/1524 der Kommission vom 11. Oktober 2018 zur Festlegung einer Überwachungsmethodik und der Modalitäten für die Berichterstattung der Mitgliedstaaten gemäß der Richtlinie (EU) 2016/2102 des Europäischen Parlaments und des Rates über den barrierefreien Zugang zu den Websites und mobilen Anwendungen öffentlicher Stellen.*

Europäisches Institut für Telekommunikationsnormen. (2018). *En 301 549 v2.1.2 - accessibility requirements for ict products and services.*

Europäisches Institut für Telekommunikationsnormen. (2020). *Draft - en 301 549 v3.2.1 - accessibility requirements for ict products and services.*

Europäisches Parlament und Rat der europäischen Union. (2016). *Richtlinie (EU) 2016/2102 des europäischen Parlaments und des Rates vom 26. Oktober 2016 über den barrierefreien Zugang zu den Websites und mobilen Anwendungen öffentlicher Stellen.*

Freire, A. P. (2012). *Disabled people and the web: User-based measurement of accessibility* (Unveröffentlichte Dissertation). University of York.

Graap, F., Haase, S., Jäckel, S., Steil, H., Voigt, H., Zehendner, E. & Zimmer, F. (2018). Freie Werkzeuge zur Beurteilung des Barrierefreiheitsgrades von Webseiten. In R. Dachselt & G. Weber (Hrsg.), *Mensch und Computer 2018 - Workshopband.* Bonn: Gesellschaft für Informatik e.V. doi: 10.18420/muc2018-ws01-0443

Heuristischer Walkthrough. (2018). https://medien-wiki.de/heuristischer-walkthrough. (zuletzt abgerufen am 19.04.2021)

Hogrefe AG. (2016). *Item.* https://dorsch.hogrefe.com/stichwort/item. (zuletzt abgerufen am 21.04.2021)

Holder, F. (2020). *Accessibility guidelines 3.0.* https://myaccessible.website/accessibility-expert/ag-3.0. (zuletzt abgerufen am 08.04.2021)

ISO. (1999, Januar). *Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten Teil 11: Anforderungen an die Gebrauchstauglichkeit - Leitsätze* (Norm Nr. DIN EN ISO 9241-11).

Janson, M. (2019). *Das Sterben der Bankfilialen.* https://de.statista.com/infografik/14350/anzahl-der-kreditinstitute-und-bankfilialen-in-deutschland/. (zuletzt abgerufen am 05.01.2021)

Lang, T. (2004). Comparing website accessibility evaluation methods and learnings from usability evaluation methods..

Mankoff, J., Fait, H. & Tran, T. (2005). Is your web page accessible? a comparative study of methods for assessing web page accessibility for the blind. In *Proceedings of the sigchi conference on human factors in computing systems* (S. 41–50). New York, NY, USA: Association for Computing Machinery. Zugriff auf https://doi.org/10.1145/1054972.1054979 doi: 10.1145/1054972.1054979

Mateen, S., Voß-Nakkour, S. & Rustemeier, L. (2020). Eine Studie zur Qualitätsbeurteilung von automatisierten Testwerkzeugen zur Prüfung auf Barrierefreiheit. In R. Zender, D. Ifenthaler, T. Leonhardt & C. Schumacher (Hrsg.), *Delfi 2020 – die 18. fachtagung bildungstechnologien der gesellschaft für informatik e.v.* (S. 157-168). Bonn: Gesellschaft für Informatik e.V.

Onlinemarketing-Praxis. (o.D.a). *Definition Webseite.* https://www.onlinemarketing-praxis.de/glossar/webseite. (zuletzt abgerufen am 19.04.2021)

Onlinemarketing-Praxis. (o.D.b). *Definition Website.* https://www.onlinemarketing-praxis.de/glossar/website. (zuletzt abgerufen am 16.04.2021)

Rømen, D. & Svanæs, D. (2008). Evaluating web site accessibility: Validating the wai guidelines through usability testing with disabled users. In *Proceedings of the 5th nordic conference on human-computer interaction: Building bridges* (S. 535–538). New York, NY, USA: Association for Computing Machinery. Zugriff auf https://doi.org/10.1145/1463160.1463238 doi: 10.1145/1463160.1463238

Schmidt, V.-M. (2013). *Verbesserung der Barrierefreiheit von Websites durch manuelle und automatische Testwerkzeuge : Handlungsempfehlungen für Bibliotheken* (Bachelor Thesis). Zugriff auf http://nbn-resolving.de/urn:nbn:de:hbz:79pbc-opus-8543

Schmutz, S., Sonderegger, A. & Sauer, J. (2016). Implementing recommendations from web accessibility guidelines: Would they also provide benefits to nondisabled users. *Human Factors*, *58* (4), 611-629. Zugriff auf https://doi.org/10.1177/0018720816640962 (PMID: 27044605) doi: 10.1177/0018720816640962

Sozialverbund VdK Deutschland e.V. (o.D.). *Grad der Behinderung (GdB).* https://www.vdk.de/deutschland/pages/themen/artikel/9216/grad_der_behinderung_gdb. (zuletzt abgerufen am 19.01.2021)

Statista. (2021a). *Marktanteile der führenden Betriebssysteme in Deutschland von Januar 2009 bis März 2021.* https://de.statista.com/statistik/daten/studie/158102/umfrage/marktanteile-von-betriebssystemen-in-deutschland-seit-2009/. (zuletzt abgerufen am 08.04.2021)

Statista. (2021b). *Marktanteile der führenden Browserfamilien an der Internetnutzung in Deutschland von Januar 2009 bis März 2021.* https://de.statista.com/statistik/daten/studie/13007/umfrage/marktanteile-der-browser-bei-der-internetnutzung-in-deutschland-seit-2009/. (zuletzt abgerufen am 08.04.2021)

Statista. (2021c). *Marktanteile der führenden mobilen Betriebssysteme an der Internetnutzung mit Mobiltelefonen in Deutschland von Januar 2009 bis März 2021.* https://de.statista.com/statistik/daten/studie/184332/umfrage/marktanteil-der-mobilen-betriebssysteme-in-deutschland-seit-2009/.

Statista. (2021d). *Marktanteile der führenden mobilen Browser an der Internetnutzung mit Mobiltelefonen in Deutschland von Januar 2009 bis März 2021.* https://de.statista.com/statistik/daten/studie/184297/umfrage/marktanteile-mobiler-browser-bei-der-internetnutzung-in-deutschland-seit-2009/. (zuletzt abgerufen am 08.04.2021)

Statistische Bundesamt. (2020). *Pressemitteilung Nr. 230 vom 24. Juni 2020.* https://www.destatis.de/DE/Presse/Pressemitteilungen/2020/06/PD20_230_227.html. (zuletzt abgerufen am 05.01.2021)

T-Systems Multimedia Solutions GmbH. (2019, 11). *Beispielbericht Barrierefreiheit.* (Internes Dokument eingesehen am 14.10.2020)

T-Systems Multimedia Solutions GmbH. (2020a, 04). *BITV-Prüfmethodik - DAkkS akkreditierter Testprozess Expertenevaluation.* (Internes Dokument eingesehen am 14.10.2020)

T-Systems Multimedia Solutions GmbH. (2020b, 06). *Verfahrensbeschreibung.* (Internes Dokument eingesehen am 14.10.2020)

T-Systems Multimedia Solutions GmbH. (2021a, 02). *Beispiel-Prüfbericht Barrierefreiheit.* (Internes Dokument eingesehen am 26.02.2021)

T-Systems Multimedia Solutions GmbH. (2021b, 03). *Konformitätsbescheinigung zur Barrierefreiheit.* (Internes Dokument eingesehen am 12.04.2021)

T-Systems Multimedia Solutions GmbH. (o.D.a). *Audio- & Video Content - Prüfen zeitbasierter Medien.* (Internes Dokument eingesehen am 26.02.2021)

T-Systems Multimedia Solutions GmbH. (o.D.b). *Bookmarklets.* (Internes Dokument eingesehen am 04.03.2021)

T-Systems Multimedia Solutions GmbH. (o.D.c). *Checkliste Barrierefreiheit Details.* (Internes Prüftool eingesehen am 26.12.2020)

T-Systems Multimedia Solutions GmbH. (o.D.d). *Dokument zur Bestimmung des Aufwands.* (Internes Dokument eingesehen am 26.02.2021)

T-Systems Multimedia Solutions GmbH. (o.D.e). *Vorlage Statusdokument Barrierefreiheit.* (Internes Prüftool eingesehen am 26.12.2020)

Velleman, E. & van der Geest, T. (2013). Page sample size in web accessibility testing: How many pages is enough? In *Proceedings of the 15th international acm sigaccess conference on computers and accessibility.* New York, NY, USA: Association for Computing Machinery. Zugriff auf https://doi.org/10.1145/2513383.2513408 doi: 10.1145/2513383.2513408

Verband der Ersatzkassen e.V. (2020a). *Daten zum Gesundheitswesen: Soziale Pflegeversicherung (SPV).* https://www.vdek.com/presse/daten/f_pflegeversicherung.html. (zuletzt abgerufen am 05.01.2021)

Verband der Ersatzkassen e.V. (2020b). *Über den Verband der Ersatzkassen e. V. (vdek).* https://www.vdek.com/ueber_uns.html. (zuletzt abgerufen am 16.04.2021)

Viehweger, J. (2009). *Untersuchung und Evaluation von Bibliothekswebangeboten unter dem besonderen Aspekt der Barrierefreiheit anhand ausgewählter Beispiele* (Diplomarbeit, Humboldt-Universität zu Berlin, null). doi: http://dx.doi.org/10.18452/2003

Vigo, M., Brown, J. & Conway, V. (2013). Benchmarking web accessibility evaluation tools: measuring the harm of sole reliance on automated tests. In *Proceedings of the 10th international cross-disciplinary conference on web accessibility* (S. 1–10).

von Weizsäcker, R. (1987). *Weihnachtsansprache 1987 von Bundespräsident Richard von Weizsäcker.* https://www.bundespraesident.de/SharedDocs/Reden/DE/Richard-von-Weizsaecker/Reden/1987/12/19871224_Rede.html. (zuletzt abgerufen am 03.03.2021)

W3C. (2015a). *Authoring tool accessibility guidelines (atag) 2.0.* https://www.w3.org/TR/ATAG20/. (zuletzt abgerufen am 21.04.2021)

W3C. (2015b). *User agent accessibility guidelines (uaag) 2.0.* https://www.w3.org/TR/UAAG20/. (zuletzt abgerufen am 21.04.2021)

W3C. (2016). *Web accessibility evaluation tools list.* https://www.w3.org/WAI/ER/

tools/. (zuletzt abgerufen am 30.03.2021)

W3C. (2018). *Web content accessibility guidelines (wcag) 2.1.* https://www.w3.org/TR/WCAG21/. (zuletzt abgerufen am 27.10.2020)

W3C. (2020). *Web content accessibility guidelines (wcag) overview.* https://www.w3.org/WAI/standards-guidelines/wcag/. (zuletzt abgerufen am 25.11.2020)

W3C. (2021). *W3c editor's draft 22 march 2021 - w3c accessibility guidelines (wcag) 3.0.* https://w3c.github.io/silver/guidelines. (zuletzt abgerufen am 08.04.2021)

W3C. (o.D.a). *About w3c.* https://www.w3.org/Consortium/. (zuletzt abgerufen am 19.04.2021)

W3C. (o.D.b). *Introduction to understanding wcag 2.1.* https://www.w3.org/WAI/WCAG21/Understanding/intro. (zuletzt abgerufen am 26.11.2020)

Williams, R. & Brownlow, S. (2020). *The click-away pound report 2019.* http://www.clickawaypound.com/downloads/cap19final0502.pdf. (zuletzt abgerufen am 09.02.2021)

Yesilada, Y., Brajnik, G. & Harper, S. (2009). How much does expertise matter? a barrier walkthrough study with experts and non-experts. In *Proceedings of the 11th international acm sigaccess conference on computers and accessibility* (S. 203–210). New York, NY, USA: Association for Computing Machinery. Zugriff auf https://doi.org/10.1145/1639642.1639678 doi: 10.1145/1639642.1639678

Zhang, M., Wang, C., Bu, J., Yu, Z., Lu, Y., Zhang, R. & Chen, C. (2015). An optimal sampling method for web accessibility quantitative metric. In *Proceedings of the 12th international web for all conference.* New York, NY, USA: Association for Computing Machinery. Zugriff auf https://doi.org/10.1145/2745555.2746663 doi: 10.1145/2745555.2746663

Zimmermann, G. (2005). Usability und Barrierefreiheit – Gemeinsam sind wir stark? In M. Hassenzahl & M. Peissner (Hrsg.), *Tagungsband up05* (S. 38-45). Stuttgart: Fraunhofer Verlag.