

Kalle Gustafsson

VERKKOPALVELUIDEN SAAVUTETTA- VUUDEN ARVIOINTI: AUTOMAATTIS- TEN TYÖKALUJEN TEHOKKUUS JA KÄYTTÖ

Informaatioteknologian ja viestinnän tiedekunta
Kandidaattitutkielma
Toukokuu 2023

TIIVISTELMÄ

Kalle Gustafsson: Verkkopalveluiden saavutettavuuden arviointi: automaattisten työkalujen tehokkuus ja käyttö
Kandidaattitutkielma
Tampereen yliopisto
Tietojenkäsittelytieteiden tutkinto-ohjelma
Toukokuu 2023

Digitalisaation myötä ihmisille tärkeät palvelut, kuten pankit, virastot ja työnhakupalvelut ovat siirtyneet verkkoon. Sen vuoksi on varmistettava, että mahdollisimman monella on mahdollisuus käyttää verkkopalveluita tasavertaisesti ja ilman esteitä. Esimerkiksi eri tavalla rajoitteiset henkilöt, kuten näkövammaiset, kuulovammaiset tai ikääntyneet voivat kokea haasteita sekä verkkopalveluiden käytössä että niihin pääsyssä. Tämän takia on tärkeää, että verkkosivut ovat saavutettavia erilaisille käyttäjille. Saavutettavuudella tarkoitetaan sitä, että pyritään vastaamaan eri tavoin rajoitteisten henkilöiden verkkopalveluiden käytön haasteisiin.

Suomessa on tullut voimaan vuonna 2019 Euroopan unionin saavutettavuusdirektiiviin pohjautuva digipalvelulaki, jolla on pyritty varmistamaan tietty saavutettavuuden taso. Laki määrittelee, että julkisen sektorin digitaalisten palveluiden on ylitettävä tietty saavutettavuuden vähimmäistaso. Tällöin saavutettavuudella tulisi taata erilaisille kansalaisille mahdollisuus käyttää digitaalisia palveluita, esimerkiksi Kelan verkkopalveluita. Saavutettavuuden merkityksen korostuminen on kannustanut kehittämään työkaluja saavutettavuuden arviointiin.

Tämän tutkielman tavoitteena on selvittää, kuinka tarkkoja tuloksia saavutettavuuden arviointityökalut antavat sekä millaisia saavutettavuusongelmia ne kykenevät tai eivät kykene tunnistamaan. Tutkielma on muodoltaan kirjallisuuskatsaus, joka on jaettu kahteen osaan. Ensimmäisessä osassa käydään läpi yleisesti saavutettavuuden määritelmä, WCAG 2.0-ohjeistus ja saavutettavuuden arviointikeinot. Toisessa osassa paneudutaan tarkemmin työkaluihin ja niiden antamien tulosten arviointiin.

Saavutettavuuden arviointityökalujen antamia tuloksia tutkimalla huomattiin, että työkalut tunnistavat varsin heikosti saavutettavuusongelmia. Ne kuitenkin auttavat verkkopalveluiden kehittäjiä varmistamaan saavutettavuuden vähimmäistason. Jotta tuloksista saadaan mahdollisimman luotettavia, tulisi kehittäjien käyttää vähintään kolmea tai useampaa työkalua arvioinnissa. Tutkimus osoittaa, että ihmisen tekemä arviointi on edelleen välttämätöntä saavutettavuuden varmistamiseksi ja tarkin tulos saadaan usein yhdistelemällä kaikkia saatavilla olevia arviointikeinoja.

Avainsanat: saavutettavuus, käytettävyys, verkkopalvelut, arviointityökalut

Tämän julkaisun alkuperäisyys on tarkastettu Turnitin Originality Check -ohjelmalla.

Sisällysluettelo

1	JOHDANTO	3
2	SAAVUTETTAVUUDESTA YLEISESTI.....	5
2.1	SAAVUTETTAVUUDEN MÄÄRITELMÄ.....	5
2.2	VERKON SAAVUTETTAVUUDEN OHJEISTUS.....	6
2.3	VERKON SAAVUTETTAVUUDEN ARVIOINTI.....	8
2.3.1	<i>Käyttäjättestaus</i>	9
2.3.2	<i>Manuaalinen arviointi</i>	9
2.3.3	<i>Automaattinen arviointi</i>	10
3	AUTOMAATTISET SAAVUTETTAVUUSARVIOINTI TYÖKALUT JA NIIDEN ANTAMIEN TULOSTEN ARVIOINTI.....	12
3.1	AUTOMAATTISET ARVIOINTITYÖKALUT	12
3.2	TYÖKALUJEN ANTAMAT TULOKSET JA NIIDEN ARVIOINTI.....	13
3.2.1	<i>Kattavuus</i>	14
3.2.2	<i>Täydellisyys</i>	15
3.2.3	<i>Virheettömyys</i>	15
3.2.4	<i>Yksityiskohtaisuus</i>	16
4	YHTEENVETO	17
	LÄHDELUETTELO.....	18

1 Johdanto

Digitalisaatio pakottaa lähes jokaisen ihmisen siirtymään sähköisten palvelujen käyttäjiksi. Useimmille elintärkeät palvelut, kuten pankit, virastot ja työnhaku ovat siirtyneet verkkoon. Tämän vuoksi on varmistettava, että verkkopalveluita pystyy käyttämään tasavertaisesti mahdollisimman suuri joukko ihmisiä. Toimenpiteet, joilla madalletaan käyttöesteitä ja helpotetaan kaikkien pääsyä verkkopalveluihin kuvastavat saavutettavuuden ydintä. Saavutettavuudella pyritään vastaamaan erilaisten käyttäjäryhmien, kuten näkö- ja kuulovammaisten sekä seniorien, erityistarpeisiin digitaalisessa ympäristössä. Näiden tarpeiden tunnistaminen on johtanut kehitystyöhön, jossa on syntynyt automatisoituja työkaluja verkkopalveluiden saavutettavuuden arvioimiseksi ja parantamiseksi. Työkaluja oli rekisteröity vuonna 2014 neljäkymmentäviisi ja vuoteen 2017 mennessä niitä oli jo yhdeksänkymmentäkolme, ja niitä kehitetään jatkuvasti lisää (Abduganiev, 2017). Vaikka saavutettavuuden merkitys ymmärretään, Lopes et al. (2010) osoittaa kuitenkin tutkimuksessaan, että suurin osa verkkosivuista ei ole saavutettavia. Syynä tähän ei ole verkon kehittäjien tietämättömyys saavutettavuudesta, vaan Vigo et al. (2013) uskoo sen johtuvan liiallisesta tukeutumisesta automaattisiin saavutettavuus arviointityökaluihin.

Tämän tutkielman tavoitteena on selvittää, kuinka tarkkoja tuloksia automaattiset työkalut antavat, millaisia saavutettavuusongelmia ne kykenevät tunnistamaan ja mitä ne eivät pysty tunnistamaan. Työssä tutkitaan ainoastaan työkaluja, jotka pystyvät antamaan itsenäisesti arvioita saavutettavuudesta. Huomioon ei oteta muita arvioinnissa käytettäviä työkaluja, kuten näytönlukijoita. Automaattisia työkaluja koskevia tutkimuksia on rajattu niin, että ne ovat korkeintaan kymmenen vuotta vanhoja, jotta saadaan tuloksia työkaluista, jotka testaavat WCAG 2.0 ohjeistusta vastaan. Lähdemateriaalina käytetty ainoastaan suomen- tai englanninkielisiä lähteitä.

Mielenkiintoni aiheeseen heräsi vierailuluennonjohtajan, sokean ohjelmistokehittäjän, ansiosta, kun hän esitteli digitaalisen saavutettavuuden merkitystä ja suositteli saavutettavuuden arviointityökalun lataamista selaimelle. Työkalun käyttö luennolla verkkosivustojen saavutettavuuden testaamiseen herätti pohdintaa työkalujen kyvystä havaita saavutettavuusongelmia ja niiden antamien tulosten paikkansapitävyydestä.

Tutkielma rakentuu siten, että ensin kappaleessa kaksi kerrotaan yleisesti saavutettavuudesta. Kappaleessa kolme kerrotaan tarkemmin saavutettavuuden arviointi työkaluista, sekä tutkimuksien tuloksista, joissa arvioidaan työkaluja. Neljännessä kappaleessa on yhteenveto aiheesta.

Tutkimus on toteutettu kirjallisuuskatsauksena. Aineistoina on käytetty tieteellisiä artikkeleita, verkkosivuja, kirjoja sekä tutkimuksia. Tietokantoja ja hakupalveluita, joita on käytetty tutkimuksessa ovat Google Scholar, ACM Digital Library, SpringerLink ja Andor. Suurin osa tutkimuksessa käytetystä materiaalista on löytynyt ACM:sta. Hakulausekkeet ovat olleet pääosin englanniksi, sillä suomeksi aineistoa löytyi hyvin rajallisesti. Suomeksi hakulausekkeet ovat olleet seuraavanlaisia: ”saavutettavuus”, ”saavutettavuuden arviointi”, ”verkon saavutettavuus”. Englanninkieliset hakulauseet on muodostettu seuraavien sanojen ”web”, ”accessibility”, ”evaluation”, ”testing”, ”tool” yhdistelmillä. Hakusanoja on yhdistelty myös käyttämällä AND ja OR operaattoreita erilaisten tulosten löytämiseksi. Englanniksi aineistoa haettaessa oli tärkeää sisällyttää ”web” jokaiseen hakulausekkeeseen, sillä ”accessibility” tarkoittaa englanniksi saavuttavuuden lisäksi myös fyysiseen maailmaan sijoittuvaa esteettömyyttä.

Englanniksi tuloksia löytyi hakulausekkeiden avulla tietokannoista satoja. Työkaluja koskevia tuloksia on rajattu valitsemalla 2010–2023 välillä julkaistut artikkelit, jotta saadaan ainoastaan relevanttia WCAG 2.0-ohjeistusta koskevaa aineistoa. Muuten valittua aineistoa on rajattu otsikoiden, sekä abstraktien perusteella.

2 Saavutettavuudesta yleisesti

Jotta voimme arvioida automaattisia saavutettavuuden arviointityökaluja, on syytä ensin määritellä tarkemmin mitä tarkoitetaan saavutettavuudella ja miten saavutettavuutta mitataan. Ensimmäisessä luvussa käymme läpi yleisesti saavutettavuuden määritelmää. Toisessa luvussa käymme läpi saavutettavuuteen WCAG-ohjeistusta, sillä suurin osa automaattisista työkaluista käyttää arvioinnissaan WCAG-kriteerejä. Kolmannessa luvussa käymme läpi erilaisia keinoja testata verkon saavutettavuutta.

2.1 Saavutettavuuden määritelmä

Saavutettavuudella pyritään puuttumaan niihin syrjiviin tekijöihin, jotka estävät rajoitteisten henkilöiden tasavertaisen käyttäjäkokemuksen (Henry, Abou-Zahra & White, 2010). Yleisesti voidaan sanoa, että saavutettavuus tarkoittaa sitä, että kuka tahansa käyttäjä pystyy riippumattaan taloudellisesta tilanteesta, maantieteellisestä sijainnista tai fyysisistä ominaisuuksista huolimatta vierailemaan missä tahansa osaa verkkosivua, ymmärtää sen sisältö, sekä olla vuorovaikutuksessa sivun kanssa. Olettaen, että käyttäjällä on käytössään tunnettu verkkoselain, käyttöjärjestelmä sekä laite (Abduganiev, 2017). Verkkoa voidaan pitää saavutettavana, kun eri tavoin rajoitteiset henkilöt voivat tasavertaisesti havaita, ymmärtää, navigoida ja vuorovaikuttaa verkossa. Tämä tarkoittaa myös sitä, että rajoitteinen henkilö voi tuottaa ja jakaa omaa sisältöä verkkoon (Henry, Abou-Zahra & White, 2010).

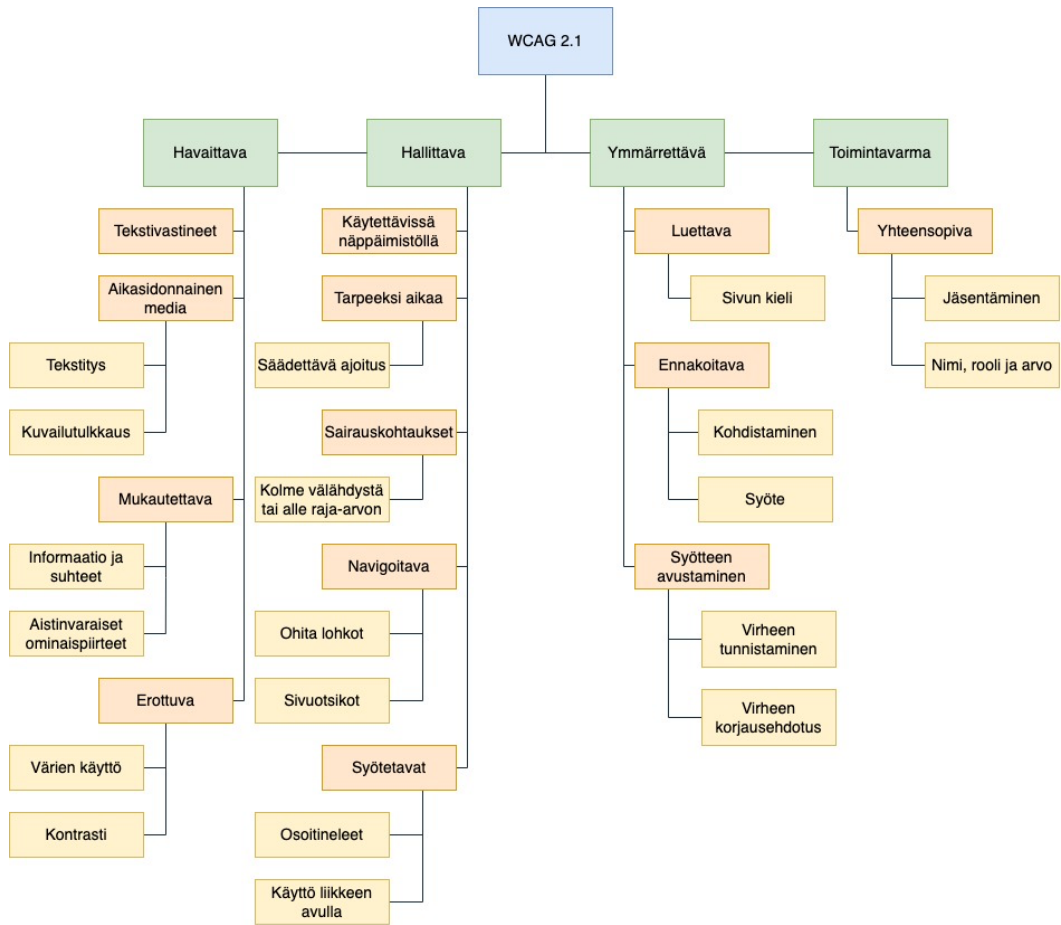
Verkon saavutettavuus pyrkii ottamaan huomioon kaikki mahdolliset rajoitteet, kuten kuulo-, kognitiiviset-, neurologiset-, fyysiset-, puhe- sekä näkörajoitteet. On hyvä huomioida, että saavutettavuus hyödyttää myös henkilöitä, joilla ei ole rajoitteita. Esimerkiksi kirkaassa valossa näyttöä selatessa, tilanteessa, jossa ei voi kuunnella ääntä tai erilaisia laitteita käytettäessä, joissa näyttöjen koot, sekä syötteenantotavat ovat erilaisia, kuten esimerkiksi puhelimissa tai älykelloissa. (Gilbert, 2019)

On tärkeää myös erottaa termit saavutettavuus (accessibility) sekä esteettömyys toisistaan. Molempia termejä on käytetty paljon ristiin, ja niiden merkitys onkin monitulkintainen. Saavutettavuudella on voitu viitata esimerkiksi tekniseen saavutettavuuteen, kuten tietoliikenneyhteyksien toimivuuteen. Esteettömyydellä puolestaan on voitu tarkoittaa niin konkreettisten kuin symbolisten esteiden poistamista. Nykyään termejä voidaan pitää vakiintuneena niin, että esteettömyydellä tarkoitetaan rakennetun ympäristön esteettömyyttä ja saavutettavuudella viitataan puolestaan sähköisiin palveluihin. (Kovanen, 2018)

2.2 Verkon saavutettavuuden ohjeistus

World Wide Web Consortium (W3C) on kansainvälinen yhteisö, jonka tehtävä on määrittellä avoimia standardeja verkon kestävä kehityksen turvaamiseksi. W3C:n Web Accessibility Initiative (WAI) on määritellyt Web Content Accessibility Guidelines (WCAG) - ohjeistuksen, jonka tehtävänä on varmistaa verkon sisällön sekä verkkosivujen saavutettavuus. WCAG-ohjeistus keskittyy erityisesti verkon tekniseen saavutettavuuteen, jolloin varmistetaan, että esimerkiksi näkövammaisten käyttäjien ruudunlukuohjelmat tai muu avustava teknologia pystyy tulkitsemaan sivun sisältöä oikein ja että käyttäjä pystyy hyödyntämään verkkopalvelun toiminnallisuuksia. WCAG-ohjeistus ei ole täydellinen, vaan sen tehtävänä on varmistaa saavutettavuuden minimitaso. Vaikka useat eri organisaatiot, kuten RNIB (Royal National Institute of Blind People) tai IBM, tarjoavat ohjeita saavutettavuuteen, pidetään verkon saavutettavuuden yleisenä ohjenuorana W3C:n WCAG-ohjeistusta, koska se on ohjeistuksista kokonaisvaltaisimmin ja kattaa muiden ohjeistusten pääpiirteet (Harper & Yesilada, 2008). WCAG-ohjeistusta käytetään myöskin monien maiden lainsäädännön saavutettavuusvaatimusten perustana.

Ensimmäinen versio W3C:n WCAG-ohjeistuksesta julkaistiin vuonna 1999. Sitä päivitettiin versioon WCAG 2.0 vuonna 2008. Uusin versio ohjeistuksesta on WCAG 2.1, jonka W3C hyväksyi kesäkuussa 2018. Nykyisen ohjeistuksen rakenne on jaettu neljän pääperiaatteen alle, jotka muodostavat verkon saavutettavuuden perustan. Periaatteita on havaittavuus, hallittavuus, ymmärrettävyys ja toimintavarmuus. Kuvassa 1 esitellään ohjeistuksen pääperiaatteet, ohjeet sekä esimerkkejä A- ja AA-tason kriteereistä.



Kuva 1. Pääperiaatteet sisältävät WCAG 2.1. ohjeistuksessa 13 yleisluonteista ohjetta. Näiden ohjeiden alle on jaettu yksityiskohtaisemmat kriteerit, joita on yhteensä 78 kappaletta.

WCAG jakaa kriteerit kolmeen eri vaatimustasoon A(matalin), AA sekä AAA(korkein):

- A-taso on vaatimustasoista alhaisin, ja se parantaa saavutettavuutta vain osalle rajoittuneista käyttäjistä. Esimerkiksi videon tekstitys vaatimus on A-tason kriteeri.
- AA-tason kriteerit parantavat saavutettavuutta laajemmalle joukolle rajoittuneita käyttäjiä. Esimerkiksi kuvailutulkkauksen tarjoaminen videoille on AA-tason kriteeri.
- AAA-tason kriteerit parantavat saavutettavuutta vielä entistä laajemmalle joukolle. AAA-tason kriteereitä on esimerkiksi sisällön tarjoamista viittomakielisenä videona, sekä sisällön kognitiivisen ymmärrettävyyden parantaminen.

Jotta verkkosivuston katsotaan noudattavan WCAG 2.1:ä on sen täytettävä kaikki kyseisen ohjeidenmukaisuustason vaatimuksista. Esimerkiksi jotta verkkosivun katsotaan olevan ohjeidenmukaisuustason AAA-tasoinen, on sen täytettävä kaikki AAA-, AA- sekä A-tason kriteerit. Suomessa vuonna 2019 voimaan tullut, Euroopan unionin saavutettavuusdirektiiviin pohjautuva, Digipalvelulaki (Finlex, 2019) määrittelee, että julkisen sektorin digitaalisten palveluiden on oltava ohjeidenmukaisuustason AA-tasoisia, jolloin ne täyttävät kaikki AA- sekä A-tason kriteerit. Digipalvelulakia sovelletaan julkisen sektorin lisäksi myös:

- Digitaalisiin palveluihin, joissa rahoituksesta vähintään puolet tulevat julkiselta sektorilta.
- Vahvoihin sähköisiin tunnistuspalveluihin.
- Vesi- ja energiahuollon, liikenteen ja postipalvelujen alan palvelua tarjoavien digitaalisiin palveluihin.
- Laissa määriteltyjen maksulaitosten, luottolaitosten, sijoituspalveluyritysten ja vakuutusyhtiöiden digitaalisiin palveluihin.

2.3 Verkon saavutettavuuden arviointi

Verkon saavutettavuuden arviointi on monilta näkökulmiltaan samanlaista, kuin ohjelmistojen laaduntarkastus. Suurimpana erona on, että ohjelmisto pysyy julkaisun jälkeen lähes muuttumattomana, kun taas verkkosivun sisältö saattaa muuttua useasti. Tämän takia verkon saavutettavuutta on arvioitava koko sen elinkaaren ajan, eikä vain tuotantovaiheessa (Abou-Zahra, 2008). Arviointiprosessissa mitataan kuinka hyvin rajoitteiset henkilöt voivat käyttää verkkoa ja voivatko rajoitteiset henkilöt käyttää verkkoa yhtä tehokkaasti, kuin henkilöt ilman rajoitteita (Abduganiev & Gaibullojonovich, 2017). Saavutettavuuden arviointi on yhdistelmä teknisiä aspekteja, kuten yleisten standardien ja ohjeistusten noudattamisen arviointia, ja ei-teknisiä aspekteja, kuten loppukäyttäjän osallistamista arviointiprosessiin. Koska verkon saavutettavuuden arviointi on kvalitatiivista eikä kvantitatiivista, on arviointi menetelmien oltava monipuolisia sekä joustavia ja niiden on sopeuduttava erilaisiin tilanteisiin (Abou-Zahra, 2008). Erilaisia testausmenetelmiä on useita. Abou-Zahra (2008) jakaa ne kolmeen ryhmään, joita ovat: automaattinen testaus, manuaalinen testaus ja käyttäjättestaus. Optimaaliset tulokset saavutettavuuden arvioinnissa saadaan, kun käytetään useita eri arviointimenetelmiä, jolloin päästään hyödyntämään kunkin menetelmän tuomia etuja. (Frazão & Duarte, 2020).

2.3.1 Käyttäjättestaus

Käyttäjättestaus on prosessi, jossa suoritetaan joko muodollisia tai epämuodollisia kokeita oikeiden käyttäjien kanssa. Testattavia pyydetään yksilöllisesti suorittamaan tavoitteen tonta tai tavoitteellista navigointia verkkosivustolla, ja asiantuntijat tarkkailevat heidän käytöstään. Lisäksi arvioijat keräävät lisätietoja käyttäjien mielipiteiden, haastattelujen ja/tai kyselyiden vastausten perusteella ja tuottavat luettelon saavutettavuusongelmista (Brajnik, Yesilada & Harper, 2011). Koska testaukseen osallistuu todellisia, vammaan omaavia, käyttäjiä, se on menetelmänä tarkempi kuin mikään approksimaatio vammasta tai psyykkisestä tilasta. Se kuitenkin vaatii paljon työtä etukäteen testin suunnitteluun ja sopivien testattavien rekrytointiin (Bai et al., 2016). Käyttäjättestauksen suurimmaksi eduksi Nuñez et al. (2019) toteaa, että tietyt saavutettavuus ongelmat sekä rikkomukset voidaan löytää suoraan, koska koetehtävät voidaan suunnitella halutun lopputuloksen löytämiseksi.

Huonoja puolia käyttäjättestauksessa Brajnikin (2008) mukaan on tehottomuus ja kyvyttömyys korostaa vikoja. Ongelmat saattavat jäädä myös huomiotta, jos ennalta määritellyt skenaariot eivät ole hyvin valittuja ja se ei tunnista matalan vakavuuden ongelmia. Lisäksi on melko monimutkaista järjestää käyttäjien testausistunto vammaisten osallistujien kanssa. Myös käyttäjättestauksen tulokset ovat todennäköisesti joukko käytettävyyso ngelmia, jotka ovat yleisiä kaikille verkkosivuston käyttäjille, eikä vain vammaisille. Toisin sanoen menetelmä todennäköisesti tunnistaa joukon todellisia, mutta merkitykset tömiä ongelmia. Tämän lisäksi, kokemattomat testien suorittajat todennäköisesti tuottavat epämääräisiä tuloksia ja pahimmassa tapauksessa jopa mitätöivät testin.

2.3.2 Manuaalinen arviointi

Käytännössä suurin osa testeistä tarvitsee edelleen manuaalista ihmisarviointia, vaikka niitä joskus ohjaavat tai tukevat ohjelmistotyökalut. Vaikka ohjelmistotyökalut voivat esimerkiksi nopeasti määrittää alt-attribuuttien olemassaolon html img -elementeissä, ihmisarvioijien on arvioitava näiden määritteiden tekstin sopivuus. Osan manuaalisista testeistä voi suorittaa ei-tekniset arvioijat, kun taas toiset testit saattavat tarvita enemmän teknistä tietämystä. Vaikka vaadittavat taidot määräytyvät pääasiassa suoritettavien testien luonteen mukaan, myös arvioijan ohjelmistotyökalujen tuki voi olla tärkeä tekijä (Abou-Zahra, 2008).

Vaativimmat asiantuntijatestit tekee henkilö, jolla on tietoa siitä, miten henkilöt, joilla on jokin vamma käyttävät verkkoa ja jotka osaavat tunnistaa käyttäjän vuorovaikutukseen liittyvät ongelmat (Abou-Zahra, 2008). Yleinen lähestymistapa asiantuntijatesteissa on tehdä heuristinen arviointi, jossa asiantuntija suorittaa käyttöliittymän arvioinnin hyväksytyjen saavutettavuuden heuristiikan tai periaatteiden perusteella. Manuaalinen WCAG-tarkastus on tästä esimerkki. Se on yksinkertainen ja tehokas menetelmä, mutta sitä on kritisoitu siitä, ettei se ole luotettava ongelmien tunnistamisessa, vaikka käytettäisiin samoja heuristiikka suosituksia (Bai et al., 2016). Brajnikin (2008) mukaan heuristisen arvioinnin etuja on kyky tunnistaa melko suuri skaala erilaisia ongelmia erilaisille käyttäjille. Se on suhteellisen kustannustehokasta, etenkin kun se suoritetaan yhdessä automaattisten testaustyökalujen kanssa. Haittapuolena on, että se vaatii taitavia arvioijia, ja että se ei todennäköisesti pysty luotettavasti erottamaan tärkeitä ja toisarvoisia saavutettavuusongelmia.

On myös mahdollista käyttää persoonakävelyä tai persoonatestausta, jossa asiantuntija simuloi tai näyttölee persoonaa tehtäviä suorittaessaan. Mitä enemmän persoonaa näyttävällä asiantuntijalla on tietoa vammaisten kohtaamista haasteista, sitä helpompi on tehdä realistinen ja uskottava arvio ratkaisusta. Tämä lähestymistapa on epämuodollinen ja nopea tehdä, mutta se on riippuvainen valituista persoonista ja asiantuntijoiden kokemuksesta. Kognitiivinen läpikulku on toinen tarkastustekniikka, jossa käyttäjien henkistä prosessia tutkitaan sen arvioimiseksi, onko käyttöliittymässä riittävästi ohjeita käyttäjien henkisen tilan tukemiseksi tehtävää suoritettaessa. Tämä vaatii toimivan prototyypin, ja jokainen vaihe on dokumentoitava huolellisesti ja luokiteltava ongelman vakavuuden määrittämiseksi. Menetelmä on aikaa vievä ja työläs, koska kaikki näkökohdat on otettava huomioon (Bai et al., 2016). ”Estekävely”-menetelmällä arvioija käyttää ennalta määritettyjen esteiden luetteloita, käyttäjäprofiilia ja mahdollista skenaariota määrittääkseen, mitä esteitä on olemassa, ja arvioidakseen niiden vakavuutta (Brajnik, Yesilada & Harper, 2011).

2.3.3 Automaattinen arviointi

Automaattisissa testeissä arvioija käyttää automatisoitua saavutettavuustyökalua, joka tarkistaa, että verkkosivu vastaa työkaluun ennalta määritettyjä saavutettavuusperiaatteita. Vaikka näitä työkaluja käytetään laajasti, ne eivät pysty täysin tunnistamaan kaikkia esteettömyysongelmia. Tämä johtuu siitä, että monet asiat edellyttävät ihmisen tekemää lisäarviointia. Työkalu voi esimerkiksi tunnistaa, onko kuvalla olemassa vaihtoehtoja tekstiä, mutta se ei osaa kertoa, onko annettu teksti sopiva annettuun kontekstiin

(Brajnik, Yesilada & Harper, 2011). Automaattinen testaus on kustannustehokasta ja voidaan suorittaa ajoittain suurille verkkosivumäärille. Testaus koskee kuitenkin vain osajoukkoa useimpien standardien mukaisista saavutettavuussäännöksistä. Tämä liittyy luonnostaan säännösten luonteeseen, jotka ovat yleensä laadullisia ja koskevat käyttöliittymää, vuorovaikutusta sekä luonnollisen kielen näkökohtia (Abou-Zahra, 2008). Saavutettavuuden testaamiseen on useita automaattisia työkaluja, jotka voivat olla esimerkiksi selainten laajennuksia, verkkosivuja tai erillisiä ohjelmistoja. Vaikka automaattisten työkalujen käyttämiseen ei tarvita syvää asiantuntijuutta, tarvitaan tietämystä kuitenkin työkalujen antamien tulosten tulkitsemiseen ja tulosten vaikutuksesta saavutettavuuteen (Bai et al., 2019).

	Menetelmät	Vahvuudet	Heikkoudet
Käyttäjätestaus	<ul style="list-style-type: none"> - Kokeet todellisten käyttäjien kanssa 	<ul style="list-style-type: none"> - Saatu informaatio tarkempaa, kuin arvio vammasta tai psyykkisestä tilasta - Tehtävät suunnitellaan itse 	<ul style="list-style-type: none"> - Työläs suunnitella ja toteuttaa - Tehoton - Lopputuloksena voi olla saavutettavuusongelmien sijaan joukko käytettävyyso ongelmia
Manuaalinen arviointi	<ul style="list-style-type: none"> - Heuristinen arviointi - Persoonakävelyt - Persoonatestaus - Kognitiivinen läpikulku 	<ul style="list-style-type: none"> - Suhteellisen kustannustehokasta - Voidaan tunnistaa suuri skaala erilaisia saavutettavuusongelmia 	<ul style="list-style-type: none"> - Lopputulokseen vaikuttaa paljon arvioijan osaaminen - aikaa vievää ja työlästä
Automaattinen arviointi	<ul style="list-style-type: none"> - Verkkosivun automaattinen arviointi työkalulla 	<ul style="list-style-type: none"> - Kustannustehokasta - Voidaan arvioida suuria määriä verkkosivuja nopeasti - Ei tarvitse syvää asiantuntijuutta 	<ul style="list-style-type: none"> - Ei tunnista kaikkia ongelmia - Käy läpi vain osan saavutettavuussäännöksistä

Taulukko 1 Tiivistelmä arviointimenetelmien vahvuuksista ja heikkouksista

3 Automaattiset saavutettavuusarviointi työkalut ja niiden antamien tulosten arviointi

Edellisessä luvussa kävimme lyhyesti läpi saavutettavuuden automaattista arviointia yleisellä tasolla. Tässä luvussa paneudumme tarkemmin arvioinnissa käytettäviin työkaluihin, sekä arvioimme niiden antamia tuloksia. Ensimmäisessä luvussa käymme läpi tutkimuksissa testatut työkalut. Toisessa luvussa paneudumme työkalujen antamiin tuloksiin ja niiden arviointiin.

3.1 Automaattiset arviointityökalut

Kuten tutkimuksessa on aiemmin jo mainittu, on erilaisia rekisteröityjä automaattisia arviointi työkaluja todella paljon. Vuonna 2017 niitä oli Abduganievin (2017) mukaan yli yhdeksänkymmentäkolme ja niitä kehitetään jatkuvasti lisää. Nykyisin työkalujen määrää mitataan jo sadoissa (W3C, 2023). Tutkimusaineistoon on valikoitunut suosituimmat ja yleisimmin käytössä olevat työkalut. Taulukossa 2 on listattuna kaikki aineistossa esiintyneet työkalut, joita on yhteensä 14 kappaletta. Macagoklun & Pekerin (2022) mukaan tällä hetkellä käytetyimmät työkalut ovat Cantan Groupin kehittämä AChecker, joka toimii omana verkkosivunaan, sekä WebAIM.org:n kehittämä WAVE-selainlaajennus. Kuvassa 2 on esimerkki WAVE-selainlaajennuksen antamista tuloksista Tampereen yliopiston verkkosivuilla.

Työkalu	Kehittäjä	Ympäristö	Lisenssi
@axe-core/cli	Deque Systems	komentorivi	ilmainen
ACCESSAssistant-Community	Level Access	selainlaajennus	ilmainen
Accessibility Insights	Microsoft	selainlaajennus	ilmainen
accessibility-checker	IBM Accessibility	komentorivi	ilmainen
AChecker	Cantan Group	verkkosivu	ilmainen
AMP	Level Access	työpöytä	maksullinen
ARC toolkit	The Paciello Group	selainlaajennus	ilmainen
aXeChromePlugin	DequeSystems	selainlaajennus	ilmainen
Cynthia Says	Cryptzone	verkkosivu	ilmainen
Deque	Deque Systems	verkkosivu	ilmainen/maksullinen
Lighthouse - accessibility	Google	selainlaajennus	ilmainen
pal ly-ci	NaturePublishingGroup	komentorivi	ilmainen
SortSite	PowerMapper	työpöytä/verkkosivu	maksullinen

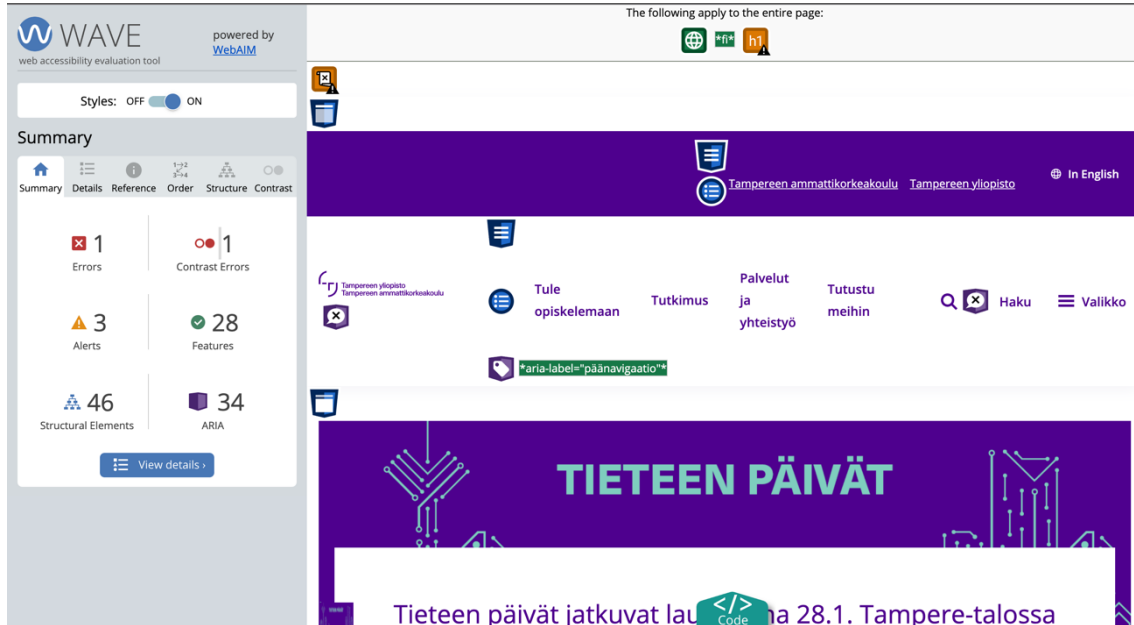
TAW	CTIC Technology Centre	verkkosivu	ilmainen
TenonCheck	Tenon LLC	selainlaajennus	ilmainen
the-all-machine	Liip	komentorivi	ilmainen
TotalValidator	Total Validator	selainlaajennus	ilmainen
WAVE Chrome Extension	WebAIM.org	selainlaajennus	ilmainen

Taulukko 2. Lähdemateriaalissa testatut saavutettavuustyökalut

3.2 Työkalujen antamat tulokset ja niiden arviointi

Jotta työkalujen tuloksia voidaan arvioida, on niitä tutkittava useasta eri näkökulmasta. Brajnik (2004) on tutkimuksessaan määritellyt kolme peruseriaatetta työkalujen tarkasteluun. Nämä periaatteet ovat englanniksi completeness, correctness sekä specificity eli vapaasti suomennettuna täydellisyys, virheettömyys sekä yksityiskohtaisuus. Täydellisyys ja virheettömyys ovat molemmat tärkeitä mittareita arvioinnissa, sillä täydellinen, mutta virheellisiä tuloksia antava työkalu saattaisi ilmoittaa jokaisella testatulla sivulla virheitä, joita välttämättä ei edes ole. Virheellisesti ilmoitettuja saavutettavuusongelmia kutsutaan false positive-tuloksiksi. Toisaalta kokonaisvaltaisen vastakohtainen työkalu, joka antaa oikeita tuloksia, mutta jättää huomioimatta saavutettavuusongelmia on myös erittäin ongelmallinen. Tällaista huomaamatta jäänyttä saavutettavuusongelmaa kutsutaan false negative-tulokseksi. False positive-tulokset aiheuttavat kohinaa tuloksiin ja aiheuttavat lähinnä häiriötä ja ärsytystä työkalun käyttäjälle. Jotta false positive-tuloksilta vältytään, voi käyttäjä esimerkiksi työkalun asetuksista muuttaa mitä saavutettavuuskriteereitä työkalu testaa. False negative-tulos on taas ongelmallisempi, sillä tällöin työkalu ei ilmoita ollenkaan potentiaalisesta saavutettavuusongelmasta. Niiden ongelmana on myös se, että työkalun käyttäjä ei voi vaikuttaa false negative-tulosten esiintymiseen. Oikeita löydettyjä ongelmia kutsutaan true positive-tuloksiksi. Tämän lisäksi työkalun antaman tuloksen yksityiskohtaisuus on huomioitava. Sillä tarkoitetaan, kuinka yksityiskohtaisesti ja tarkasti työkalu osaa kuvata potentiaalista ongelmaa. Mikäli työkalu antaa vain yleisluonteisia ohjeita löydettyistä ongelmista, ilman tarkempaa selitystä, vaikeuttaa se huomattavasti kehittäjän kykyä ymmärtää ja korjata löydetty ongelma. (Brajnik, 2004). Jotta on pystytty arvioimaan, onko työkalujen antamat tulokset true positive-, false positive- tai false negative-tuloksia on asiantuntija arvioinut manuaalisesti testatun verkkosivuston vertailua varten. Edellä esiteltyjä näkökulmia käytetään yleisesti muissakin tutki-

muksissa, mutta niiden lisäksi Vigo et al. (2013) sekä Abduganiev (2017) nostaa työkalujen kattavuutta, englanniksi coverage, koskevan näkökulman. Paneudumme kattavuuteen tarkemmin seuraavassa luvussa.



Kuva 2. WAVE-työkalun käyttöliittymä ja tulokset Tampereen yliopiston verkkosivuston saavutettavuudesta.

3.2.1 Kattavuus

Kattavuudella tarkoitetaan, kuinka hyvin työkalut tunnistavat WCAG-ohjeistuksessa määritellyjä kriteerejä. Yksi työkalujen perusongelmista onkin kuinka heikosti ne kattavat WCAG 2.0-ohjeistuksen kriteerejä, sekä kuinka jokainen työkalu käyttää eri kriteereitä jopa samojen sivujen arvioimiseen (Abduganiev, 2017). Abduganievin (2017) mukaan WAVE, joka oli heikoiten suoriutunut testatuista työkaluista, kattoi ainoastaan 12.7 % WCAG-kriteereistä. Parhaiten testatuista suoriutui Cynthia Says, joka kattoi 32.4 % 71:stä kriteeristä. Keskiarvollisesti työkalut kattoivat Abduganievin (2017) tutkimuksessa vain 20 % WCAG-kriteereistä. Vigo et al. (2013) on tutkimuksessaan samoilla linjoilla, sillä heidän mukaansa työkalut kattavat 20–50 % kriteereistä. TAW oli työkaluista ainut, joka kattoi puolet kriteereistä.

Kun analysoidaan, miten työkalut kattavat kriteereitä eri WCAG-ohjeistuksen pääperiaatteiden mukaan huomataan, että tuloksissa on merkittäviä eroja. Työkalut kattoivat parhaiten havaittavuuteen vaikuttavat rikkomukset. Havaittavuuden jälkeen seuraavaksi rikotuin pääperiaate on toimintavarmuus (Abduganiev, 2017). On tietysti selvää, että on rajoitteita, mitä voidaan automaattisilla työkaluilla testata ja mitä ei. Esimerkiksi hallittavuuden pääperiaatteen alle kuuluvan kriteerin 2.1.2 ”Ei näppäimistöänsää” testaaminen

vaatii oikeaa vuorovaikutusta tai simulaatiota. On kuitenkin kriteerejä, jotka olisivat testattavissa tarkistamalla esimerkiksi tietyn HTML elementtien olemassaolo tai sisältö, mutta työkalut eivät silti sitä tee. Esimerkiksi osa työkaluista ei testaa onko verkkosivulla otsikkoa, joka on WCAG-ohjeistuksen kriteeri 2.4.2 ”Sivuotsikot. (Vigo et al., 2013).

3.2.2 Täydellisyys

Täydellisyydellä mitataan, onko työkalujen tunnistamat saavutettavuusrikkomukset todellisia rikkomuksia, kun tuloksia verrataan asiantuntijan tekemään arviointiin. Toisin sanoen täydellisyys ilmaisee, kuinka hyvin työkalu minimoi false negative-tulokset ja maksimoi true positive-tulokset. (Vigo et al., 2013). Työkalut eivät tarjoa niin hyvää täydellisyyttä, kuin voisi olettaa (Abduganiev, 2017). Abduganievin mukaan paras työkalu, Cynthia Says, tuotti 59 % true positive – tuloksia. Parhaimmassa tapauksessa työkalut tunnistavat keskiarvollisesti neljä virhettä yhdeksästä. Tutkittaessa millaisia saavutettavuusrikkomuksia työkalut tunnistavat kaikkein täydellisimmin huomataan, että siihen vaikuttaa WCAG-kriteeristön vaatimustaso (A, AA ja AAA), sekä minkä pääperiaatteen alle rikkomus kuuluu. Työkalut tunnistavat täydellisimmin toimintavarmuus pääperiaatteen liittyvät rikkomukset, sekä matalan vaatimustason (A) rikkomukset. Työkalujen tulokset ovat sitä epätäydellisempiä, mitä korkeammalle WCAG-kriteeristön vaatimustasolle mennään. (Vigo et al., 2013). Syyksi tähän Vigo et al. epäilee sen johtuvan siitä, että työkalujen kehittäjät priorisoivat A-tason rikkomuksia. Toisaalta voidaan myös olettaa, että mitä korkeamman saavutettavuustason rikkomus on kyseessä, sitä vaikeampi se on tunnistaa. Vigo et al. korostaa myös, että työkalujen täydellisyyttä mitattaessa tuloksiin vaikuttaa myös millä tasolla testattavan sivuston saavutettavuus on lähtökohtaisesti. Mitä huonommin saavutettavaa sivua testataan, sitä täydellisempiä tuloksia työkalut antavat. Työkalujen antamien tuloksien erot siis pienenevät, mitä huonommin saavutettavaa sivua testataan. Syynä tähän on todennäköisesti se, että korkean saavutettavuustason sivustoilla rikkomukset ovat hienovaraisia ja vaikeita tunnistaa. Toisaalta matalan saavutettavuustason sivustoilla esiintyvät rikkomukset ovat yleisiä sekä odotettavissa olevia. Voidaan siis sanoa, että työkalut ovat suunniteltu huomaamaan stereotyyppisiä sekä usein ilmeneviä saavutettavuusongelmia. (Vigo et al., 2013)

3.2.3 Virheettömyys

Virheettömyys, josta voidaan käyttää myös termiä tarkkuus, mittaa kuinka monta prosenttia työkalujen ilmoittamista ongelmista ovat todellisia. Voidaan siis sanoa, että se viittaa siihen, kuinka usein työkalun ilmoittama tulos on false positive ja siten minimoi

niiden määrää (Abduganiev, 2017). Yleisesti työkalut osoittavat korkeaa virheettömyyttä, joka on yli 93 % (Vigo et al., 2013). Lisäksi on huomattu, että virheettömimmin toimivat työkalut ovat epätäydellisimpiä. Täydellisyydellä ja virheettömyydellä on siis tärkeä syy-yhteys ja niillä on keskeinen rooli tehokkuuden karakterisoinnissa. (Abduganiev, 2017).

3.2.4 Yksityiskohtaisuus

Yksityiskohtaisuudella mitataan työkalun kykyä tarjota ylimääräistä ja erottuvaa tietoa löytyneistä todellisista saavutettavuusongelmista, joita muut työkalut eivät pysty ilmaisemaan. Se huomioi, miten työkalu selittää ongelman ja kuinka se ohjaa ja neuvoo kehittäjää ongelman ratkaisemisessa. Mahdollisimman tarkka yksityiskohtaisuus on olennainen osa työkalun hyödyllisyyttä, sillä se auttaa kehittäjiä ymmärtämään ongelmia paremmin. Abduganievin (2017) tutkimuksen mukaan Cynthia Says -työkalu tarjosi parhaan yksityiskohtaisuustason, kun taas SortSite, TAW ja Tenon jäivät jälkeen tässä suhteessa, mikä voi heikentää niiden käytännön hyödyllisyyttä ongelmien korjaamisessa.

4 Yhteenveto

Tässä tutkielmassa tutkittiin automaattisten saavutettavuuden arviointityökalujen tehokkuutta verkkopalveluiden saavutettavuuden arvioinnissa. Tutkimuksen lähtökohtana on ollut selvittää, kuinka tarkasti ja kattavasti nämä työkalut pystyvät tunnistamaan ja raportoimaan erilaisia saavutettavuusongelmia verkkosivustoilla ja sovelluksissa.

Tutkielmassa käydään aluksi läpi saavutettavuuden peruskäsitteitä, selitetään miksi verkkopalveluiden saavutettavuus on tärkeää, ja mitä se käytännössä tarkoittaa. Tutkielmassa esitellään saavutettavuuden määritelmiä ja ohjeistuksia, kuten WCAG 2.0 -standardi. Työssä esitellään useita suosittuja saavutettavuuden arviointityökaluja, kuten WAVE, AXE ja Lighthouse, ja tarkastellaan niiden kykyä tunnistaa erilaisia saavutettavuusongelmia. Tutkielmassa analysoidaan myös, kuinka nämä työkalut vertautuvat toisiinsa ja millaisia vahvuuksia ja heikkouksia kullakin on.

Tutkielman keskeinen havainto on, että vaikka automaattiset arviointityökalut ovat hyödyllisiä ja tarjoavat nopean tavan tunnistaa joitakin ilmeisiä saavutettavuusongelmia, ne eivät pysty havaitsemaan kaikkia mahdollisia ongelmia. Työkalut ovat parhaimmillaan saavutettavuuden perustason varmistamisessa, mutta esimerkiksi monimutkaisemmat saavutettavuusongelmat ja kontekstisidonnaiset virheet jäävät usein tunnistamatta. Tämän vuoksi tutkielmassa korostetaan, että näiden työkalujen käyttö yksin ei ole riittävä keino taata verkkopalvelun täydellistä saavutettavuutta, vaan ihmisen tekemä arviointi on edelleen tarpeellista. Parhaat tulokset saavutettavuuden arvioinnissa saadaankin, kun yhdistetään useita eri arviointimenetelmiä. Tämä sama pätee myös arviointityökalujen käyttöön, sillä mitä useampaa eri työkalua käyttää, sitä tarkempia tulokset ovat. Suositeltavaa on tehdä arvio vähintään kolmella eri työkalulla.

Lähdeluettelo

- Abduganiev, S. G., 2017. Towards automated web accessibility evaluation: a comparative study. *Ijites* 9, 18–44. doi:10.5815/ijites.2017.09.03
- Abou-Zahra, S. 2008. Web Accessibility Evaluation. In: Harper, S., Yesilada, Y. (eds) *Web Accessibility. Human-Computer Interaction Series*. Springer, London. https://doi-org.libproxy.tuni.fi/10.1007/978-1-84800-050-6_7
- Bai, A., Mork, H. C., Schulz, T. & Fuglerud, K. S. 2016. Evaluation of Accessibility Testing Methods. Which Methods Uncover What Type of Problems? 506-516. doi: 10.3233/978-1-61499-684-2-506
- Brajnik, G. 2004. Comparing accessibility evaluation tools: a method for tool effectiveness. *Universal Access in the Information Society* 3, 252–263. <https://doi.org/10.1007/s10209-004-0105-y>
- Brajnik, G. 2008. A comparative test of web accessibility evaluation methods. In *Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility (Assets '08)*. Association for Computing Machinery, New York, NY, USA, 113–120. <https://doi-org.libproxy.tuni.fi/10.1145/1414471.1414494>
- Brajnik, G., Yesilada, Y. & Harper, S. 2011. The Expertise Effect on Web Accessibility Evaluation Methods, *Human-Computer Interaction*, 26:3, 246-283, DOI: 10.1080/07370024.2011.601670
- Digipalvelulaki (2019). Laki digitaalisten palvelujen tarjoamisesta 306/2019. FINLEX. <https://www.finlex.fi/fi/laki/alkup/2019/20190306> (viitattu 8.8.2022)
- Frazão, T. & Duarte, D. 2020. Comparing accessibility evaluation plug-ins. In *Proceedings of the 17th International Web for All Conference (W4A '20)*. Association for Computing Machinery, New York, NY, USA, Article 20, 1–11. <https://doi-org.libproxy.tuni.fi/10.1145/3371300.3383346>
- Gilbert, R. M. 2019. Inclusive Design for a Digital World. https://doi.org/10.1007/978-1-4842-5016-7_2
- Kovanen, K. (2018). Verkkopalveluiden saavutettavuuden huomiointi: Finna.fi - palvelun saavutettavuustestaus näkövammaisille käyttäjille. Pro Gradu -tutkielma. Tampereen yliopisto. Viestintätieteiden tiedekunta. Informaatiotutkimus ja interaktiivinen media.
- Lopes, R., Gomes, D. & Carriço, L. 2010. Web not for all: a large scale study of web accessibility. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A) (W4A '10)*. Association for Computing Machinery, New York, NY, USA, Article 10, 1–4. <https://doi-org.libproxy.tuni.fi/10.1145/1805986.1806001>

Macagoklu, Ş. S. and Peker, S. 2022. "Web accessibility performance analysis using web content accessibility guidelines and automated tools: a systematic literature review," 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Ankara, Turkey, pp. 1-8, doi: 10.1109/HORA55278.2022.9799981.

Nuñez, A., Moquillaza, A., & Paz, F. 2019. Web Accessibility Evaluation Methods: A Systematic Review. In: Marcus, A., Wang, W. (eds) Design, User Experience, and Usability. Practice and Case Studies. Lecture Notes in Computer Science, vol 11586. Springer, Cham. https://doi-org.libproxy.tuni.fi/10.1007/978-3-030-23535-2_17

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 (W4A '13). Association for Computing Machinery, New York, NY, USA, Article 1, 1–10. <https://doi.org/10.1145/2461121.2461124>

W3C. Web Accessibility Evaluation Tools List. 2023. <https://www.w3.org/WAI/ER/tools/> (viitattu 19.5.2023)