

TEKNILLINEN KORKEAKOULU

Elektroniikan, tietoliikenteen ja automaation tiedekunta

Ilari Pulkkinen

Käytettävyytutkimuksen yksiselitteisyys

Kandidaatin työ
Espoo 10.12.2008

Työn ohjaaja:

Dosentti Timo Korhonen

Tekijä: Ilari Pulkkinen		
Työn nimi: Käytettävyyystutkimuksen yksiselitteisyys		
Päivämäärä: 10.12.2008	Kieli: Suomi	Sivumäärä: 3 + 32
Tutkinto-ohjelma: Tietoliikennetekniikka		
Vastuuopettaja: TkT Riku Jäntti		
Ohjaaja: Dosentti Timo Korhonen		
<p>Tämän työn tavoitteena on ottaa esille keskeisimmät käytettävyyystutkimuksen ongelmat ja esittää ehdotuksia näiden ongelmien ratkaisemiseksi. Yksi suurimmista käytettävyyystutkimusmenetelmien ongelmista on se, että niiden tuottamat tulokset eivät ole riittävän yksiselitteisiä ja näin tulosten vertailu on todella haastavaa.</p> <p>Työn alkupuolella esitellään useita tunnetuimpia käytettävyyystutkimusmenetelmiä ja nostetaan esille niiden heikkouksia ja vahvuuksia. Tämän jälkeen kerrotaan kuinka eri menetelmiä tulisi yhdistää, jotta saataisiin tehtyä riittävän kattava käytettävyyystutkimus.</p> <p>Tämä työ perustuu hyvin pitkälti käytettävyystudkija Rolf Molichin ja kumppaneiden laajoihin ammattimaisten käytettävyysslaboratorioiden suorittamiin käytettävyyssarvioihin saman tuotteen suhteen. Näissä tutkimuksissa selviää, että eri käytettävyyssasiantuntijat löytävät todella erilaisia käytettävyyso ongelmia, vaikka he käyttäisivät samoja tehtäviä käytettävyystestissä. Sen lisäksi tutkimus paljastaa, että palveluista, joita pidetään yleisesti käytettävyydeltään hyvinä löydetään helposti useita satoja käytettävyyso ongelmia.</p> <p>Työn loppupuoli keskittyy pohtimaan vastauksia näihin ongelmiin ottamalla esille mahdollisia ratkaisuvaihtoehtoja. Tässä työssä selvitetään, mikä on riittävä käyttäjien ja arvioijien määrä käytettävyystestissä. Mitä tulisi tehdä, jotta käytettävyystudkimusmenetelmistä saataisiin systemaattisempia. Mitä tulisi tehdä, jotta löydettyistä käytettävyyso ongelmista olisi enemmän hyötyä kuin mitä tällä hetkellä on.</p>		
Avainsanat: käytettävyys, käytettävyystudkimus, käytettävyystestaus, käytettävyyssarvio		

Sisällysluettelo

Tiivistelmä.....	ii
Sisällysluettelo.....	iii
1 Johdanto.....	1
2 Käytettävyytutkimukseen kohdistuvaa kritiikkiä.....	2
3 Tutkimusaineisto ja menetelmät.....	4
3.1 Tapaustutkimukset, joiden pohjalle työ rakentuu.....	4
3.2 Käytettävyytutkimuksen menetelmät.....	6
3.2.1 Käytettävyyttestaus.....	6
3.2.2 Ääneen ajattelu -menetelmä (Think-Aloud).....	7
3.2.3 Heuristinen arvio.....	8
3.2.4 Näkökulmaan perustuva arvio (Perspective-based Usability inspection).....	10
3.2.5 Käyttäjätiedon automaattinen kerääminen webissä.....	11
3.2.6 Kognitiivinen läpikäynti (Cognitive Walkthrough).....	12
3.3 Käytettävyytutkimusmenetelmien vertailu.....	12
3.4 Heuristisen arvion ja näkökulmaan perustuvan arvion vertailu.....	13
3.5 Heuristisen arvion, käytettävyyttestauksen ja automaattisen käyttäjätiedon keräyksen vertailu.....	13
3.6 Kognitiivisen läpikäynnin ja heuristisen arvion vertailu.....	14
3.7 Laboratoriotestauksen ja kenttätestauksen vertailu.....	14
3.8 Äänen ajattelu -menetelmän ja kognitiivisen läpikäynnin vertailu.....	15
3.9 Yhteenveto vertailusta.....	15
3.10 Kuinka käytettävyytutkimuksen menetelmiä tulisi käyttää yhdessä, jotta ne paikkaisivat toistensa puutteita.....	16
4 Johtopäätökset.....	18
4.1 Mitä voidaan tehdä, jotta käytettävyytutkimusmenetelmillä saataisiin parempia tuloksia.....	18
4.2 Käyttäjien ja käytettävyyttestaajien määrän merkitys.....	18
4.3 Tarvitaan ongelmien löytämiseen lisää järjestelmällisyyttä.....	19
4.4 Tarvitaan tehokkaampia keinoja luokitella ja karsia vähäpätöisiä ongelmia.....	23
5 Yhteenveto.....	27
Lähteet.....	28

1 Johdanto

Jo 1980-luvulla Käytettävyysasiantuntija Jakob Nielsen [1] sanoi, että ohjelmistot ja tuotteet sisältävät liikaa ominaisuuksia, enemmän kuin kukaan tulee koskaan tarvitsemaan. Tämä lisäksi Allwoodin [2] 1980-luvulla tekemän kenttätutkimuksen mukaan ihmiset tuhlaavat 5-10 prosenttia työajastaan tietokoneiden käytöstä johtuviin vaikeuksiin. Näin ollen kun laitteiden ja ohjelmistojen ominaisuudet ja määrä kasvavat, niin varsinaiseen työhön käytetty aika vähenee yhä entisestään.

Yksi tärkeimmistä asioista, jolla voidaan parantaa tuotteen käyttäjätuottavuutta on sen on käytettävyys. SFS-EN ISO 9241-11 -standardin [3] mukaan käytettävyys mittaa, miten hyvin käyttäjät voivat käyttää tuotetta tietyssä käyttötilanteessa saavuttaakseen määritetyt tavoitteet tuloksellisesti, tehokkaasti ja tyytyväisinä.

Tietojenkäsittelytieteen alkuaikoina tuotteiden ja palveluiden käytettävyyteen ei panostettu. Nykyään varsinkin IT-teknologian aloilla tuotteita ei osteta, joiden käytettävyys on heikko. Käytettävyyttä pidetään markkinavalttina [4]. Kristine Faulkner [2] mainitsee kirjassaan myös, että kun tiedostetaan täsmälleen mitä käyttäjä haluaa pystytään luomaan parhaita mahdollisia tuotteita näille käyttäjille tiettyihin tehtäviin tiettyyn ympäristöön.

On olemassa lukuisia artikkeleita [5],[6],[7],[8], [9], joissa esitellään yhä uudempia ja parempia ja monimutkaisempia käytettävyystutkimusmenetelmiä. Kuitenkaan ei ole kovinkaan montaa artikkelia, joissa esitettäisiin, miten eri menetelmiä tulisi käyttää yhdessä ja miten tulisi ottaa huomioon eri käytettävyystutkimusmenetelmien puutteet.

Tämä työ on tehty vertailevana kirjallisuustutkimuksena. Tämän työn tavoitteena on selittää käytettävyystutkimuksen ongelmia ja esittää lähtökohtia näiden ongelmien ratkaisemiseksi. Yksi suuri käytettävyystutkimusmenetelmien ongelma on se, että niiden tuottamat tulokset eivät ole riittävän yksiselitteisiä ja näin tulosten vertailu on todella haastavaa. Kvantitatiivista käyttäjädataa keräävät menetelmät ovat yksiselitteisiä, mutta niiden tarjoama tieto ei kerro käyttäjien tekojen taustalla vaikuttavista asioista, ja tämä on oleellista kun etsitään käytettävyysongelmia. Näin ollen suurimmassa osassa tapauksista kvalitatiivisten menetelmien tulee toimia pääasiallisina käytettävyystestausmenetelminä.

Tämä työn alussa tarkastellaan käytettävyystutkimukseen kohdistuvaa kritiikkiä. Eniten käytössä olevien subjektiivista dataa tuottavien käytettävyysmenetelmien arviointiin ei olla vielä löydetty kunnan kriteereitä. Myös menetelmien tuottamat tulokset ovat kaikkea muuta kuin yksiselitteisiä. Tämän jälkeen esitellään Molichin ja kumppaneiden tutkimukset [10][11], jotka toimivat pohjana tälle työlle. Tässä työssä esitellään yleisimpiä käytettävyystutkimusmenetelmiä, vertaillaan niiden ominaisuuksia ja pohditaan kuinka niitä tulisi käyttää yhdessä. Lopuksi pohditaan, mitä haasteita käytettävyystutkimuksella on ja kuinka näihin haasteisiin tulee vastata.

2 Käytettävyydestutkimukseen kohdistuvaa kritiikkiä

Nielsenin työssä (1994) [12] kerrotaan, että on yleisessä tiedossa, että kirjallisuudessa esitettyjä käytettävyyden arviointimenetelmiä ei juuri käytetä käytännön käytettävyydestestauksessa osana ohjelmistojen kehittämisprojekteja. Myös Molichin ja kumppaneiden työssä [10][11] käy ilmi, että ammattimaiset käytettävyydestestausfirmat eivät käytä kirjallisuudessa suositeltuja käytettävyydenarviointimenetelmiä. Näihin vähän käytettyihin menetelmiin kuuluvat myös se, että tuotekehitysprosessin alkuvaiheissa tulisi keskittyä käyttäjään ja ottaa mukaan iteratiivinen suunnittelutapa. Nielsen [12] kertoo, että yksi tärkeä syy siihen, että kirjallisuuden suosittelamia käytettävyydenarviointimenetelmiä ei käytetä on se, että niiden toteuttaminen on kallista ja aikaa vievää. Syy siihen, miksi kirjallisuudessa esitettyjä kattavampia menetelmiä ei käytetä on, että niillä ei saavuteta riittävän paljon parempia tuloksia verrattuna halvempiin menetelmiin.

Kai Öörnin työ [13] nostaa esille useita ongelmia käytettävyyesarviointiin liittyen. Kun tehdään käytettävyyesarviota, tulee valita sopivat käytettävyyesarviointimenetelmät sitä varten. Millä perusteella tämä valinta tulee tehdä? Mitkä ovat oikeat valintakriteerit? Whittaker et al. [14] (2000) ottavat työssään esille ongelman, kuinka vaikea on valita oikeat arviointikriteerit käytettävyyesarviolle. Ongelma on se, että käytettävyyesarvioijat joutuvat yleensä itse määrittämään tuotteen tai palvelun ongelmat, vaatimukset ja käytettävyyden parantumiseen käytetyt mitta-asteikot. Hartson et al. [15] selvittävät, että alalla ei ole vielä pystytty muodostamaan selkeitä valintakriteereitä eri käytettävyydenarviointimenetelmille. Työssä todetaan myös, että koska käytettävyyesarviointimenetelmien tutkimus on sen verran nuori ala, niin vertailevia tutkimuksia ei vielä ole kovinkaan paljoa. Todetaan, että käytettävyydenarviointimenetelmät sinänsä eivät ole vakaita, sillä ihmisen ja tietokoneen vuorovaikutukseen perustuvat tuotteet ja palvelut kehittyvät todella nopeasti ja ne luovat kokoajan uusia vaatimuksia käytettävyyden arviointimenetelmille.

Vaikka valmiita arviointikriteereitä on olemassa (kuten ISO 9126), niitä ei käytetä sillä ne ovat liian yleisellä tasolla ja niiden muokkaaminen vastaamaan tiettyä kontekstia on liian työlästä, sillä arvioitavien tuotteiden kirjo on todella laaja. Myös Hartsonin et al. työ [15] tarjoaa mahdolliseksi arviointikriteereiksi läpikotaisuuden, pätevyyden ja luotettavuuden. Työssä todetaan kuitenkin, että tarvitaan lisää tutkimusta, jotta nämäkään kriteerit olisivat riittävän hyviä luotettavien tulosten saamiseksi.

Kun ollaan valittu sopivat menetelmät käytettävyyden arvioimiseen sekä päämäärät, jotka tulee arviossa saavuttaa ongelmaksi tulee, mistä tiedämme, että menetelmien avulla löydettyihin ongelmiin puuttamalla ollaan oikeastaan parannettu tuotteen käytettävyyttä? Öörnin työssä [13] esitetyn kirjallisuuskatsauksen mukaan tähän mennessä ei olla pystytty määrittämään mitään riittävän luotettavaa tapaa mitata käytettävyyden parantumista kun kvantitatiiviset menetelmät ovat käytössä. Joten ainut tapa arvioida käytettävyyttä on turvautuminen subjektiiviseen arvioon ja arvioida samankaltaisia tuotteita tai palveluita keskenään ja tällöin saadut tulokset eivät missään nimessä ole yleistettävissä [13]. Vaikka pystyttäisiinkin määrittämään kvantitatiivisesti eri menetelmien kriteerit käytettävyyden parantamiseksi silloin ongelmaksi jäisi vielä se, että onko tulos tilastollisesti uskottava [13]. Lähes aina käytettävyydestesteissä oleva otos käyttäjistä ei ole tilastollisesti merkittävä [13].

Yhtenä ongelmana voidaan pitää usein sitä, että eri käytettävyydenarviointimenetelmiä on todella paljon. Ivory & Hearts [9] listaavat 75 eri metodia. Alan tutkimuksessa ollaan liikaa keskitytty uusien käytettävyydmetodien tuottamiseen. Syvä ymmärrys siitä, mitä informaatiota eri menetelmillä saadaan ja kuinka niiden avulla voidaan parantaa käytettävyyttä, on puutteellinen.

Cordes [16] kertoo, että Suurimmassa osassa käytettävyydesteistä testin suunnittelija valitsee sellaisia tehtäviä, joita testissä oleva tuote tukee. Sen sijaan tulisi valita tehtäviä, jotka ovat oleellisia koko sille tuoteryhmälle, mitä tuote edustaa. Jos testin tekevä käyttäjä tietää tämän, sillä on todella isoja vaikutuksia käyttäjän suoritukseen testissä ja hänen asenteisiinsa tuotetta kohtaan ja näin varsinkin kvantitatiivisista tuloksista tulee hyvin epäluotettavia.

Hertzumin ja Jacobsenin tekemässä työssä [17] puolestaan selviää, että laajimmin käytössä olevissa käytettävyydenarviointimenetelmissä, kuten heuristisessa arvioissa (heuristic evaluation), ääneen ajattelu -menetelmässä (think-aloud), sekä kognitiivisessa läpikäynnissä (cognitive walkthrough) arvioijalla on suuri vääristävä vaikutus tuloksiin (näitä menetelmiä käsitellän tarkemmin 3. kappaleen menetelmät -osiossa). Eri arvioijien löytämät käytettävyysongelmat eroavat myös hyvin paljon toisistaan. Hertzumin et al. [17] työn mukaan tämä johtuu siitä arvioija ammentaa tietoa omasta kokemuksestaan ja osaamisestaan.

Kun eri käytettävyydenarviointimenetelmillä löydetään käytettävyysongelmia niiden perusteella tehdään parannusehdotuksia tuotteeseen. Lewis [18] esittelee työssään käsitteen nimeltä ongelmien löytymisaste (problem-discovery rate) (p), joka määrittää tietyn käyttäjämäärän suhteen keskimääräisen käytettävyysongelmien määrän, eri käyttäjien välillä. Lewisin tutkimuksessa huomataan, että 11 eri käytettävyydetutkimuksen perusteella ongelmien löytymisastetta liioitellaan. Käytettävyydetutkimuksessa pienellä käyttäjämäärällä tehdystä tutkimuksesta, sekä tutkimuksen subjektiivisesta luonteesta johtuen ei voi vetää johtopäätöksiä koskien suurempaa käyttäjäryhmää. Täytyy olla jokin tapa skaalata pienellä käyttäjäryhmän otoksella tehdyt arviot antamaan riittävän luotettava tulos koskien koko käyttäjäryhmää.

3 Tutkimusaineisto ja menetelmät

3.1 Tapaustutkimukset, joiden pohjalle työ rakentuu

Työssä käytetty tutkimusaineisto nojaa vahvasti Molichin ja hänen kollegoidensa julkaisemiin artikkeleihin Comparative Usability Evaluation [10] ja Comparative Usability Evaluation (CUE-4) [11].

Ensin esittelyssä Molichin ja kumppaneiden CUE-2 tutkimus [10]. Vuosina 1998 ja 1999 kahdeksan käytettävyyslaboratoriota osallistui kokeeseen. Jokaisen laboratorion tehtävänä oli itsenäisesti arvioida www.hotmail.com -palvelun käytettävyyttä. Ohjeistuksessa ilmoitettiin myös testausskenaario, jossa oli kuvattu käytettävyysarvion tavoitteet.

Tiimit saivat itse valita, mitä käytettävyysmetodeita he käyttäisivät. Osallistujia pyydettiin käyttämään sellaisia metodeja, jotka olisivat mahdollisimman lähellä heidän normaaleja käytäntöjään. Kaikki paitsi yksi tiimi käyttivät ”think-aloud” -menetelmää. Jotkut tiimit tukivat myös arviotaan asiantuntija-arvioilla ja kyselymenetelmillä. Tiimejä pyydettiin myös kertomaan, millä tavalla heidän käytettävyyden arviointitapansa erosivat tavallisesta tämän kokeen aikana, mitä resursseja he käyttivät arvioon kuten työtunnit, sekä kommentteja siitä, kuinka realistisia heidän tekemänsä arviot olivat olleet.

CUE-2 työn tarkoituksena oli tarjota katsaus siitä, miten johtavat www-sivujen käytettävyyden arvioijat toimivat. Sen yhtenä tavoitteena oli tarjota perusta käytettävyystestauksen metodologiselle ja teoreettiselle keskustelulle. Tarkoituksena oli myös näyttää mukana oleville käytettävyyslaboratorioille heidän vahvuutensa ja heikkoutensa käytettävyyden arvioinnissa hienovaraisen itsearvioinnin kautta.

Yllättävää oli se, että vaikka kaikille tiimeille annettiin Hotmailin tärkeimpien ominaisuuksien lista niin eri tiimien suunnitelmat, sekä käytettävyystestauksessa käytetyt tehtävät olivat todella erilaisia.

CU2- työssä selvisi, että eri tiimit löysivät hyvin erilaisia käytettävyysongelmia, eikä ollut sellaista yksittäistä ongelmaa, jonka kaikki tiimit olisivat löytäneet. Vielä yllättävämpää oli se, että ne tiimit jotka olivat käyttäneet käytettävyystestauksessa samanlaisia tehtäviä löysivät myös hyvin erilaisia käytettävyysongelmia. Kaiken kaikkiaan 75 % käytettävyysongelmista olivat sellaisia, jotka eivät löytyneet kahden tai useamman tiimin raportista.

Tässä työssä huomattiin myös, että käytettävyydeltään hyväksi uskotussa palvelusta kuten Hotmailista löydettiin todella paljon käytettävyysongelmia, kaikenkaikkiaan eri käytettävyystiimit löysivät 310 käytettävyysongelmaa. Olettaen, että Hotmail edustaa johtavaa esimerkkiä hyvästä web käytettävyydestä niin voimme päätellä, että käytettävyysongelmien määrä tyypillisessä www-sivussa on niin iso, että emme voi toivoa löytävämmekään kuin pienen murto-osan kaikista sen käytettävyysongelmista perinteisellä käytettävyystestillä

Nielsen väitti, että kun käytettävyystestissä käytetään viittä eri käyttäjää löydetään 85 % käytettävyysongelmista. Kuitenkin CUE-2 projektissa 9 tiimiä löysi 310 käytettävyysongelmaa, josta 85 % olisi ollut 264 ongelmaa. Kuitenkaan mikään yksittäinen tiimi ei löytänyt näin montaa käytettävyysongelmaan. Eli tässä valossa yksittäisellä käytettävyystestillä ei pystytä löytämään 85 % kaikista käytettävyysongelmista (kaikilla tarkoitetaan sitä, mitä kaikki

käytettävyyssiimit löytävät yhteensä).

Toinen Molichin ja kumppaneiden su orittama työ oli CUE-4 [11]. Siinä 17 ammattimaista käytettävyystestaustiimiä arvioivat erikseen New Yorkilaisen Hotel Pensylvanian www-sivun käytettävyyttä. Tässä tutkimuksessa yhdeksän tiimiä käyttivät perinteistä käytettävyystestausta ja loput kahdeksan käyttivät asiantuntija-arviomenetelmiä, kuten heuristista arviota (katso menetelmät, osio 3.). Myös CUE-4 tutkimuksessa oli käytössä asiakasskenaario, jonka tarkoituksena oli simuloida suhdetta arvioitavan sivun tuotekehitystiimiin ja se myös antoi yhteisen aloituspisteen käytettävyystestaustiimeille. Tämän skenaarion kirjoittivat arvioitavasta sivusta vastaavat henkilöt. Skenaario sisälsi listan, jossa oli sivuston kolme tärkeintä tehtävää ja neljä osa-aluetta, joita voisi arvioida. Molich ja kumppanit huomasivat, että edellisessä CUE-2 tutkimuksessa [10] käytettävyyssiimit olivat raportoineet liian suuren määrän käytettävyyso ongelmia, jotta niitä olisi järkevää käsitellä. Näin ollen CUE-4 tutkimukseen Molich ja kumppanit asettivat raporttiin 50 käytettävyyso ngelman rajan estääkseen tämän. Jokainen tiimi sai vapaasti valita omat käytettävyy sarviointimenetelmänsä, mutta jotkut käytettävyystestaustiimit antoivat Molichin ja kumppaneiden johtaman tutkimustiimin päättää mitä metodologia he käyttävät. Yksi tämän tutkimuksen tarkoituksista oli selvittää, mitä menetelmiä alan käytettävyyssammattilaiset oikeastaan käyttävät.

Tiimit, jotka käyttivät käytettävyystestausta suorittivat 5-15 testisessiota. Kaikki käytettävyystestausta käyttäneet tiimit hyödynsivät myös ääneen ajattelu -menetelmää (think-aloud), sekä sen tukena testauksen jälkeisen subjektiivisen kokemuksen arviointiasteikkoa (QUIS) tai kyselyä. Tiimit, jotka käyttivät asiantuntija-arvioita käyttivät joko ”perspective-based” -arviointimenetelmää, kilpailija-analyysia tai heuristista arviota tai muita vastaavia menetelmiä.

Myös CUE-4 tutkimuksessa havaittiin, että eri käytettävyyssiimit löysivät hyvin erilaisia käytettävyyso ongelmia. Syyksi tälle epäiltiin sitä, että käytettävyyso ngelmien määrä vain on niin rajattoman suuri, joten väistämättä löydetään hyvin erilaisia ongelmia. Työssä epäiltiin, että oikeat www-palvelut (toisin kuin yhden sivun prototyypit) sisältävät monia eri polkuja ja näin ollen mahdollisten käyttäjälle vapaana olevien kombinaatioiden määrä on suuri, joten myös löydettyjen käytettävyyso ngelmien määrä on todella suuri.

Kuten CUE-2 -tutkimuksessa, myös CUE-4 -tutkimuksessa käytettävyy sarviointitiimit löysivät paljon käytettävyyso ongelmia, yhteensä 340. Todettiin, että käytettävyy sasiantuntijat, jotka erikseen arvioivat tuotetta löytävät satoja käytettävyyso ngelmia jopa hyvin yksinkertaisesta tuotteesta ja moni näistä ongelmista on vakava tai jopa kriittinen. Ongelmia löydetään siis enemmän kuin niitä pystytään käsittelemään. Näin ollen työssä todetaan, että käytettävyyden arviointimenetelmät ovat liian herkkiä havaitsemaan käytettävyyso ngelmia, sillä vaikka tutkimuksessa olleesta sivustosta löytyi 340 käytettävyyso ngelmaa sitä pystyttiin silti käyttämään.

CUE-4 -tutkimuksessa selvisi, että asiantuntija-arvioista saatua dataa voidaan verrata käytettävyystesteistä saatuun dataan. Eli kun asiantuntija-arvion suorittaa kokenut käytettävyy salan ammattilainen sen tuottamat tulokset ovat pätevyydeltään vertailtavissa käytettävyystestaukseen. Työssä on myös selvästi nähtävissä, että asiantuntija-arviointimenetelmillä ja käytettävyy stestauksella löydetään erilaisia ongelmia. Toisaalta kun vertaillaan eri käytettävyy stestejä keskenään myös silloin löydetään todella erilaisia ongelmia, joten käytettävyy stestillä ei löydetä sen varmemmin käytettävyyso ngelmia kuin asiantuntija-arviollakaan. Myös CUE-2 tutkimus [10] tukee tätä havaintoa. Työssä todetaan myös, että tämä havainto on ristiriidassa aikaisemman tutkimuksen kanssa.

CUE-4 tutkimuksessa käsiteltiin myös vääriä hälytyksiä (false alarm). Väärä hälytys on ehdotettu muutos perustuen johonkin käytettävyysoongelmaan, jolla ei olisi ollenkaan positiivista vaikutusta käytettävyyteen ja joka voisi tehdä palvelusta jopa vähemmän käytettävän. Yleinen uskomus käytettävyytutkimuksen alalla on se, että asiantuntija-arviot tuottavat useita vääriä hälytyksiä kun taas käytettäjätestauksen tuloksia pidetään yleisesti luotettavampina. Jos tämä olisi totta vain asiantuntija-arviolla löytyisi todella paljon sellaisia käytettävyysoongelmia, jotka vain yksi tiimi olisi raportoinut. Kuitenkin 32.2 % kaikista käytettävyytestillä tehdyistä löydöistä raportoitiin yhden tiimin toimesta ja 27.9 % kaikista asiantuntija-arviolla tehdyistä löydöistä raportoitiin myös yhden tiimin toimesta. Näin ollen työssä todetaan, että tulokset eri menetelmillä ovat varsin samoja. Kuitenkin, koska kyseessä ei ole tilastollisesti merkittävä tulos tätä johtopäätöstä ei voida pitää täysin pitävänä. Tutkimuksessa epäillään, että syy tähän luuloon lienee se, että aikaisemmin asiantuntija-arviolla vääriä hälytyksiä tuli sen takia, että ei käytetty riittävän erilaisia käytettävyystestejä vahvistamaan asiantuntija-arviolla saatuja ongelmia. Joten käytettävyystestaus ei ole selkeästi paras menetelmä, johon muita menetelmiä tulisi verrata.

Tutkimuksessa todettiin myös, että käytettävyystestit eivät ole tehokkaita, kun palvelun suunnitellut tuotekehitystiimi ei ole mukana yhteistyössä. Kun suunnittelijat eivät ole mukana suunnittelussa ja arviointiprosessissa tai käytettävyyssarvioijilla ei ole mahdollisuutta työskennellä suunnittelijoiden kanssa mahdollisuus siihen, että tuotteeseen tulee parannuksia vähenee.

Myös Jacobsen et al. [19] tutkivat, kuinka samaan palveluun tehdyt käytettävyytutkimukset eroavat toisistaan. Tässä työssä neljä HCI tutkijaa arvioivat samat videonauhut neljästä eri käytettävyystestisessiosista, jossa käytettiin ääneen ajattelu -menetelmää. Myös tässä tutkimuksessa havaittiin, että eri käytettävyyssarvioijat löytävät hyvin erilaisia käytettävyysoongelmia. Tämä johtuu Jacobsenin mukaan siitä, kuinka arvioija itse vaikuttaa testauksessa olevan käyttäjän toimintaan tulosta vääristävästi.

Myös Kessner et al. [20] tekivät Molichin ja kumppaneiden CUE-2 työtä vastaavan työn. Siinä kuusi ammattimaista käytettävyystiimiä teki arvion samalle tuotteelle. Myös siinä selvisi, että eri tiimit löytävät todella erilaisia käytettävyysoongelmia. Yhdestä palvelusta löydettiin myös todella paljon käytettävyysoongelmia, kuten myös Molichin ja kumppaneiden [10][11] tutkimuksissa.

Tässä työssä pyritään löytämään ratkaisuvaihtoehtoja CUE-2 ja CUE-4 tutkimuksissa nousseisiin ongelmiin. Pääasiallisesti käsittelyssä on se, että eri käytettävyyssarvioijat löytävät todella erilaisia ongelmia ja se, että hyvänä pidetystäkin tuotteesta on helposti löydettävissä paljon käytettävyysoongelmia.

3.2 Käytettävyytutkimuksen menetelmät

Työssä arvioitavien menetelmien pohjana on Nielsenin [1] esittämä menetelmävalikoimaa, jolla pystytään suorittamaan käytettävyyssarvio minimaalisilla resursseilla lyhyessä ajassa. Syy tähän lähestymistapaan on siinä, että nämä menetelmät ovat yleisimmin käytössä olevat käytettävyytutkimuksen menetelmät, kuten myös Molichin ja kumppaneiden [10][11] tutkimuksessa huomataan. Koska nämä ovat yleisimpiä menetelmiä, niin tämä työ pureutuu juuri niihin ongelmiin, jotka vallitsevat ammattimaisessa käytettävyytutkimuksessa.

3.2.1 Käytettävyystestaus

Käytettävyystestauksessa [21] nähdään kuinka oikean käyttäjäryhmän edustajat käyttävät

tuotetta ennalta valittujen tehtävien suorittamiseen. Siinä käyttäjä suorittaa tehtävät joko prototyypillä tai valmiilla tuotteella samalla kun tutkija tarkkailee käyttäjän käyttäytymistä ja kerää empiristä dataa siitä, miten käyttäjä suorittaa tehtävää useimmiten videoimalla käyttötapahtuman. Tyypillisesti testauksessa mitataan käyttäjän tehtävään suorittamiseen kuluvaa aikaa, virheiden määrää ja käyttäjäytyyväisyyttä. Nämä tiedot toimivat pohjana listalle, jossa luetellaan havaittuja käytettävyyssongelmia. Havaintojen pohjalta tehdään myös parannusehdotuksia tuotteeseen.

Hyvä käytettävyydestä on huolellisesti suunniteltu ja selkeästi koordinoitu. Perusteellinen käytettävyydestä lähtee liikkeelle testin tavoitteiden määrittämisestä. Tällaisia tavoitteita voi esimerkiksi olla hyvin yleisiä, kuten tuotteen käyttäjäryhmän käyttöytyyväisyyden parantaminen ja tuotteen käytön helppouden lisääminen. Tavoitteet voivat olla myös hyvin tarkkoja, kuten esimerkiksi valikoiden muokkaaminen, jotta ne tehostaisivat käyttäjän toimintaa.

Kun tavoitteet on määritelty tulee valita otos tuotetta käyttävästä käyttäjäryhmästä. Otokseen tulevien ihmisten tulee edustaa kaikkia niitä eri ihmistyyppisiä, jotka kuuluvat tuotteen kohderyhmään. Valittavien käyttäjien määrä riippuu testin luonteesta. Molich ja kumppanit [11] ehdottavat, että jos kyseessä on esimerkiksi sairaalan mittauslaitteiston käytettävyydestä tulee ottaa tavallista enemmän käyttäjiä, jotta todennäköisyys havaita enemmän käytettävyyssongelmia kasvaisi.

Kun käyttäjät on valittu täytyy määrittää tehtäviä, joita käyttäjät tekevät testattavan tuotteen avulla. Näiden tehtävien tulee edustaa sellaisia asioita, joita käyttäjäryhmän jäsenet normaalisti tekisivät tuotteen kanssa. Tehtävätyyppien valinnassa voidaan hyödyntää tuotteen suunnitteluvaiheessa asetettuja päämääriä, jotka tuotteen tulee toteuttaa.

Ennen käytettävyydestä on syytä määrittää kriteerit, joiden avulla käytettävyydestä saatuja tuloksia arvioidaan. Kriteerit voivat olla hyvin kvalitatiivisia kuten käyttäjäytyyväisyys, käytön vaikeus tai hyvin kvantitatiivisia kuten tehtävään käytetty aika, käyttäjän tekemien virheiden määrä ja onnistuneesti suoritettujen tehtävien määrä.

Testiä suorittaessa testaaaja voi käyttäjän tarkkailun lisäksi hyödyntää eri menetelmiä, kuten ääneen ajattelu -menetelmää, jossa käyttäjää pyydetään selostamaan, mitä hän on tekemässä tai ”co-discovery” -menetelmää, jossa kaksi käyttäjää yhdessä suorittaa tehtävää auttaen toinen toistaan tai ”active intervention” -menetelmää, jossa testaaaja kysyy testin lomassa käyttäjältä asioita liittyen käyttäjän suorittamaan testiin. Vielä ennen testiä tulee valmistella testausympäristö testaukseen, sekä suunnitella testauksen kulku ja muun testauksessa auttavan henkilökunnan toiminta.

3.2.2 Ääneen ajattelu -menetelmä (Think-Aloud)

Ääneen ajattelu -menetelmän tarkoituksena on saada tietoa siitä, miten käyttäjät käyttävät palvelua ja millä strategioilla he suorittavat käytettävyydestä annettuja tehtäviä. Ääneen ajattelu -menetelmää voidaan käyttää myös missä tahansa tuotekehitysprosessin vaiheessa. [2]

Menetelmän toteutuksessa käyttäjä suorittaa etukäteen suunniteltua ja valittua tehtävää ja samalla kertoo ääneen siitä, mitä hän on tekemässä ja mitä hän ajattelee [2]. Tutkimus voidaan suorittaa esimerkiksi käytettävyysslaboratoriossa. Tämä sessio yleensä videoidaan, äänitetään tai otetaan sen tapahtumista otetaan kuvankaappauksia. Myös käyttäjän näppäinten painallukset tallennetaan. Session lopuksi kirjoitetaan raportti, jossa yhdistetään video tai kuvankaappaukset käyttäjän kommentteihin ja näppäinten painalluksiin. [22]

Tällä menetelmällä käytettävyyssiantuntija pystyy ymmärtämään paremmin, miten käyttäjä kokee tuotteen, miten hän lähestyy tuotetta ja mihin hän perustaa päätöksensä liikkeessaan palvelun käyttöliittymässä. Näin ollen on helppo huomata, jos ja miten käyttäjä ymmärtää palvelun toiminnan väärin [1]. Tällöin huomataan suoraan, mitkä käyttöliittymän ominaisuudet tai funktiot aiheuttavat eniten ongelmia, koska äänen ajattelu -menetelmässä käyttäjä tulkitsee jokaisen käyttöliittymän ominaisuuden käytön. Jos ääneen ajattelu -menetelmää käyttäessä käyttäjä kokee, että ne asiat, jotka hän joutui tekemään päästäkseen tavoitteeseensa olivat jotain avain muuta kuin mitä hän oli alun perin ajatellut, voidaan todeta, että käyttöliittymä on liian monimutkainen. [12]

Ääneen ajattelu -menetelmän heikkouksiin voidaan lukea se, että kaikki eivät ole tottuneet ajattelemaan ääneen samalla kun suorittavat tehtävää ja tämä voi vaikuttaa heidän suoritukseensa [2]. Tämän menetelmän yksi haittapuoli on myös se, että sen tukena on hyvin vaikea käyttää mitään suorituskykyä mittaavaa menetelmää. Toisaalta ääneen ajattelu -menetelmää tarjoaa hyvin paljon kvalitatiivista dataa hyvinkin pienellä käyttäjämäärällä toteutettuna. Usein myös käyttäjän kommentit sisältävät eläviä ja selkeitä kuvauksia, jotka tekevät käytettävyyssraportista ymmärrettävämmän ja luettavamman [1].

3.2.3 Heuristinen arvio

Heuristisessa arviossa yleensä useampi käytettävyyssiantuntija käy läpi palvelun ominaisuuksia ja funktiota verraten niitä valmiiksi sovittuihin yleisesti hyväksytyihin ohjenuoriin, joita kutsutaan heuristiikoiksi. Idea on siinä, että hyvä käyttöliittymä noudattaa näitä heuristiikkoja. Moni käytettävyyssiantuntija on esittänyt omia heuristiikkojaan tämän arvion tueksi. Heuristinen arvio voidaan suorittaa missä tahansa tuotekehityksen vaiheessa, tosin se on paras tehdä silloin kun vielä ei ole mitään konkreettista testattavaa [12]. Kirjallisuudessa esitettyjä heuristiikkoja ovat esimerkiksi Nielsenin 10 heuristiikkaa [23], Schneidermanin 8 kultaista sääntöä ja Normanin 7 periaatetta [2]. Esittelen tässä esimerkin omaisesti Nielsenin 10 heuristiikkaa [24, suomennokset] , jotta saadaan käsitys, mitä ne oikeastaan ovat:

1. Palvelun tilan näkyvyys

Palvelun täytyisi aina informoida käyttäjää riittävästi selkeällä ja oikea-aikaisella palautteella, jotta hän tietäisi missä tilassa palvelu on.

2. Palvelun ja tosielämän vastaavuus

Palvelun tulisi käyttää tavallisesta elämästä tuttuja termejä, sanontoja ja käsitteitä palvelun erikoistermistön sijaan. Palvelun tulisi myös noudattaa tavallisesta elämästä tuttuja käytäntöjä, jotta informaatio esitettäisiin luonnollisessa ja loogisessa järjestyksessä.

3. Käyttäjän kontrolli ja vapaus

Kun käyttäjä tekee virheen hänelle tulisi tarjota selkeä ja helppo tie palata takaisin aikaisempaan tilaan. ”peru” ja ”tee uudestaan” -komentojen käyttö on suositeltavaa. Palvelu ei myöskään saisi tehdä häiritseviä asioita kysymättä käyttäjältä lupaa tähän.

4. Yhteneväisyys ja standardit

Palvelun antaman palautteen ja toimintojen pitäisi tarkoittaa yhteneväisesti aina samoja asioita, eivätkä käsitteiden merkitys vaihtuisi lennossa. Olemassa olevia käyttöliittymä- ja muita standardeja pitäisi hyödyntää yhteneväisyyden lisäämiseksi.

5. Virheiden estäminen

Palvelun suunnittelussa tulisi ottaa huomioon mahdolliset virhetilanteet ja estää niiden synty. Palvelun tulisi tunnistaa virhetilanteet ja estää niiden toistuminen informoimalla käyttäjää ennen virheen tapahtumista.

6. Tunnistaminen mieluummin kuin muistaminen

Asioiden, toimintojen ja vaihtoehtojen tulisi olla näkyvillä käyttöliittymässä. Käyttöliittymään painikkeiden ja syötteiden tulisi liittyä palvelun toimintoihin loogisesti niin, että ne olisivat tarkoituksen mukaisia palvelun toiminnan kannalta. Käyttäjältä ei pitäisi vaatia, että hän muistaa edellisen tilan asiat siirryttäessä uuteen tilaan.

7. Käytön joustavuus ja tehokkuus

Palvelun tulisi ottaa huomioon sekä aloittelevat, että edistyneemmät käyttäjät. Edistyneemmille käyttäjille palvelun tulisi tarjota pikavalintoja, sekä mahdollisuus muokata usein käytettyjä toimintoja. Laitteisto ei saisi vaikuttaa käytön joustavuuteen ja tehokkuuteen.

8. Esteettinen ja minimalistinen design

Ruudulla pitäisi näkyä vain tieto ja toiminnot, jotka ovat oleellista palvelun kannalta ja mitä tarvitaan usein. Oleellinen tieto ja toiminnot pitäisi esittää mahdollisimman selkeästi ja ytimekkäästi.

9. Virhetilanteiden tunnistaminen, ilmoittaminen ja korjaaminen

Virheilmoitusten tuliskertoa selkokielellä ja tarkasti mitä tapahtui, miksi näin kävi ja ehdottaa, miten asia tulisi korjata.

10. Opastus ja ohjeistus

Vaikka palvelun käytön tulisi onnistua ilman opastusta ja ohjeita, niin olisi hyvä tarjota käyttäjälle mahdollisuus tutustua niihin. Ohjeiden tulisi olla helposti saatavilla, nopeasti löydettävissä, oikeita käyttötilanteita tukevia antaen askel askeleelta apua palvelun ominaisuuden käyttöön. Ohjeiden tulisi olla myös mahdollisimman lyhyitä.

Yleensä 3-5 käytettävyyssiantuntijaa (jokainen yksinään) käy käyttöliittymän yksityiskohtaisesti läpi ainakin kahdesti katsoen noudattavatko palvelun ominaisuudet

heuristiikkoja. Heuristisen arvion tueksi on hyvä olla skenaarioita palvelun käyttöön liittyen, jotta käytettävyyssiantuntijat pystyvät keskittymään oleellisiin asioihin. Arvion jälkeen tulee koota saadut tulokset. Voidaan esimerkiksi kirjoittaa raportti, jossa arvioija kertoo löydöistään tai arvioija voi arvionsa aikana kertoa jollekin toiselle ihmiselle, mitä hän löytää. Kun yksittäiset arviot on suoritettu, arvion tehneet käytettävyyssiantuntijat kokoontuvat yhteen. He muodostavat löydetyistä käytettävyysongelmista kategorioita ja sovittavat löydöksensä näihin [12]. Lisätietoa heuristisen arvion toteutuksesta löytyy Nielsenin kotisivuilta [25]. Heuristinen arvio on nopea tehdä ja se vaatii vain vähän resursseja. Se antaa hyviä ideoita käyttöjärjestelmän parantamiseen [12].

3.2.4 Näkökulmaan perustuva arvio (Perspective-based Usability inspection)

Zhijun Zhangin ja kumppaneiden työ [26] esittelee näkökulmaan perustuvan arvion (Perspective-based Usability inspection). Tämän työn mukaan eri arvioijien on hankala havaita kaikkia käytettävyyso ongelmia samanaikaisesti, ja näkökulmaan perustuva arvio tarjoaa tähän parannusta. Tässä menetelmässä useassa eri arviointisessioissa keskitytään kerrallaan vain muutamaaan käytettävyyden osa-alueeseen, joiden käytettävyyttä halutaan parantaa. Nämä käytettävyyden osa-alueet tarjoavat oman käytettävyyden näkökulmansa (perspective) asiaan, listan arvioita tukevista kysymyksistä, sekä tavan toteuttaa arvio. Ideana on, että kun keskitytään kapeaan ongelma-alueeseen tarkasti määritellyillä menettelytavoilla niin jokainen sessiolla löydetään paljon ongelmia. Kun eri sessiot yhdistetään, löydetään huomattavasti enemmän ongelmia, kuin mitä normaalilla arviointimenetelmällä saataisiin sessioiden määrän pysyessä samana.

Menetelmän toteuttamisen pohjaksi tulee valita jokin käytettävyyden malli (model of human-computer interaction). Zhang et al. [26] työssä käytettiin käytettävyyden mallina Normanin seitsemän toiminnan vaihetta (Seven Stages of Action), johon oli lisätty myös virheiden käsittely. Sen kahdeksan kohtaa sisältävät:

- 1) Muodosta päämäärä
- 2) Muodosta tarkoitus
- 3) Tunnista toiminto
- 4) Toteuta toiminto
- 5) Hahmota systeemin vastaus
- 6) Tulkitse tulokset
- 7) Ymmärrä lopputulos
- 8) Reagoi virheisiin, joita on mahdollisesti syntynyt.

Tässä mallissa päämäärä on toiminto, jolla päästää lähemmäksi tehtävän suorittamista. Esimerkiksi jos tehtävänä on ostaa WWW-palvelusta tuote, niin päämäärä, jolla päästään lähemmäksi tuotteen ostamista voi olla tuotteen valitseminen tuoteluottelosta. Kun käyttäjä käyttää palvelua hän suunnittelee, mitä toimia hänen tulee tehdä saavuttaakseen päämääränsä. Tämän jälkeen hän toteuttaa nämä toiminnot, havaitsee ja ymmärtää palvelun antaman palautteen ja arvioi tilanteen, jos jokin menee vikaan.

Käytettävyyden mallin rakentamisen jälkeen määritellään eri käytettävyyden näkökulmat joiden

puitteissa palvelua arvioidaan. Eri käytettävyyšnäkökulmien tarkoituksena on keskittää arvioijan huomio johonkin palvelun tiettyyn osa-alueeseen arviointisessioiden aikana. Nämä käytettävyyšnäkökulmat tulisi olla mahdollisimman paljon toisistaan poikkeavia, sekä niiden yhdistelmän tulisi kattaa mahdollisimman laaja skaala käytettävyyden osa-alueita. Esimerkkinä voisi olla jaottelu aloitteleviin käyttäjiin ja kokeneisiin käyttäjiin. Näin yhdessä käytettävyyden arviointisessiossa tulisi ottaa huomioon aloittelevan käyttäjän näkökulma ja toisessa kokeneen käyttäjän näkökulma.

Käytettävyyden näkökulmien määrittämisen jälkeen käydään läpi käytettävyyden malli kohta kohdalta ja mietitään, minkälaisia ongelmia käyttäjä voisi mahdollisesti kohdata. Näin saadaan joukko kysymyksiä, jotka toimivat arvion tukena jokaiselle käytettävyyden näkökulmalle erikseen.

Suunnitelman lopuksi tulee määrittää erilaisia menettelytapoja eri käytettävyyden näkökulmia hyödyntävälle arviointisessiolle. Näiden menettelytapojen tarkoituksena on auttaa arvioijia, jotta oikeat käytettävyyteen liittyvät asiat tulee käytyä läpi oikeaan aikaan, sekä vähentää mahdollisuutta, että joitakin käytettävyyden ongelmia jää kokonaan huomaamatta. Tulee myös määritellä mahdollisia tehtäväskenaarioita, joita käyttäjä voisi käydä läpi käyttäessään palvelua. Yksinkertaisuudessaan arvioijien tulee ennalta määrittää minkälaisista lähtökohdista tiettyyn näkökulmaan sopivat käyttäjät lähtisivät suorittamaan palveluun liittyviä tehtäviä. Esimerkiksi aloittelijan näkökulmassa arvioijat ajattelevat käyttäjää, joka ei tunne käyttöjärjestelmää. Näin ollen tällainen käyttäjä tarvitsee opastusta, jotta hän löytää oikeat toiminnot, opetusta toimintojen käyttämiseen, erotyistä keskittymästä, jotta hän huomioi palvelun antaman palautteen, sekä oppia ymmärtämään tekojensa seurakset. Tämän jälkeen arvioijat käyvät läpi edellä luetellut asiat ja arvioivat kuinka onnistuneesti aloittelevat käyttäjät toimivat näiden asioiden suhteen. Tässä vaiheessa aikaisemmin saadut käytettävyydsarvion tukena olevat kysymykset järjestellään edellä mainittujen käyttäjän kokemien vaiheiden mukaisesti.

Arvion alkaessa tulisi olla kasassa useita eri käytettävyyden näkökulmia, sekä kysymyksiä jokaiselle eri näkökulmalle, jotka on luotu käytettävyyden mallin pohjalta. Näin toteutettuna näkökulmaan perustuva arvio tunnistaa jos palvelu ei mahdollista käyttäjän suunnittelemaa toimintaa tai palvelu esittää asiat eri tavalla kuin käyttäjä olisi olettanut.

Myös CUE-4 -tutkimuksessa [11] mainittiin menetelmä näkökulmaan perustuva (perspective-based) arvio tai ”persona-based” -arvio, sekä todettiin, että näkökulmaan perustuva arvio löytää ongelmia, jotka liittyvät myös palvelun hyödyllisyyteen, eikä pelkästään käytettävyyteen.

3.2.5 Käyttäjätiedon automaattinen kerääminen webissä

Yksi tapa kerätä automaattisesti käyttäjätiedon Internetissä on nettisivuilla käytettävät dataa keräävät lokikirjat. Tämä voidaan suorittaa kahdella eri tavalla: joko käyttäjä itse suorittaa datan keräämisen tai datan kerääminen suoritetaan systeemin yhteyteen liitettävällä automaattisella datankeräyspalvelulla. Jos käyttäjä itse suorittaa datan keräyksen se voi vaikuttaa suuresti siihen, kuinka käyttäjä suorittaa tehtävää. Jos datan keräys suoritetaan automaattisesti systeemistä käsin saavutetaan vähemmän vääristyneitä tuloksia, sillä käyttäjä ei itse joudu puuttumaan datan keräämiseen sinänsä. Tässä menettelytavassa käyttäjän tulee kuitenkin tietää, että hänen käyttöönsä tarkkaillaan oikeusteknisiin syihin vedoten. Koska käyttäjä on tietoinen siitä, että hänestä kerätään dataa voi vaikuttaa hänen suorituskykyynsä varsinkin alussa kun hän ei ole vielä tottunut datan keräämiseen. Kuitenkin jonkin ajan päästä käyttäjä tottuu datan keräykseen ja vääristävä vaikutus vähenee. [2]

Tällä menettelytavalla saadaan kerättyä todella iso määrä dataa, joten datan käsittelykin olisi syytä hoitaa hyödyntäen automaattisia menetelmiä hyödyntäen esimerkiksi tilastollista liikenneanalyysiä, jolla saadaan selville, millä alasuivilla käyttäjät viettävät eniten aikaa ja näin voidaan päätellä jotain käyttäjien mieltymyksistä. Tällöin ei kuitenkaan saada selville juuri mitään käyttäjien käyttäytymisestä. Tästä syystä ollaan menty vielä yksi askel eteenpäin ja sen nimi on Webin käyttö -analyysi (web usage analysis), sillä se pyrkii ottamaan huomioon myös käyttäjän käyttäytymisen. [21]

Webin käyttö -analyysi (web usage analysis) on varsin uusi tutkimusalue ja se keskittyy rekonstruoimaan lokitiedostojen perusteella käyttäjien webbisivun selailutilanteen. Kuten edellä Webin käyttö -analyysissä kerätään tietoa käyttäjän liikkeistä webbisivulla ja tiedot tallennetaan serverin lokiin jonkin valmiin standardin mukaisesti. Sen lisäksi tallennetaan myös semanttista informaatiota tarkoituksena selvittää polut, joilla käyttäjät vierailevat useiten ja mistä sivuista nämä polut koostuvat.[21]

3.2.6 Kognitiivinen läpikäynti (Cognitive Walkthrough)

Kognitiivisessa läpikäynnissä käytettävyyssiantuntija suorittaa tuotteelle tai palvelulle tyypillisiä tehtäviä yrittäen imitoida normaalin tuotteen käyttäjäryhmään kuuluvan käyttäjän käytöstä ja ajatella, mitä ongelmia käyttäjä mahdollisesti voisi kohdata käyttäessään tuotetta. Eli käytettävyyssiantuntijalla on samaan aikaan kaksi roolia, käyttäjän rooli, sekä asiantuntijan rooli. Kognitiivisen läpikäynnin suorittaminen vaatii käytettävyyssiantuntijalta riittävää perehtymistä tuotteen käyttäjäryhmän käyttäytymiseen tuotetta käyttäessään. Näin ollen asiantuntijan tulisi tietää, millaisia etukäteistietoja käyttäjillä mahdollisesti on ennen kuin he aloittavat tehtävien suorittamisen tuotteella.

Tämä menetelmä on hyvin nopea ja halpa, sillä siinä ei tarvita oikeita käyttäjiä. Toisaalta se vaatii todella paljon osaamista käytettävyyssiantuntijalta:

- Kykyä ymmärtää suoritettavan tehtävän luonne,
- Taitoa jakaa tehtävän suoritus erillisiin vaiheisiin,
- Taitoa ennustaa tyypillisen käyttäjän käyttäytymistä, sekä ymmärtää käyttäjien kognitiiviset kyvyt.

Kognitiivisen arvion suorittamiseen tarvitaan kuvaus tuotteen käyttäjäryhmästä, jossa kerrotaan käyttäjien taitotaso ja käsityksiä käyttäjäryhmän erityispiirteistä. Tarvitaan tarkka kuvaus testattavan tuotteen ominaisuuksista ja niiden hyödyntämisestä. Sen lisäksi tarvitaan kuvaus tyypillisestä tehtävästä, joka tuotteella suoritetaan. Lopuksi tarvitaan vielä lista osatehtävistä, jotka johtavat koko tehtävän suorittamiseen. [2]

3.3 Käytettävyystudkimusmenetelmien vertailu

Käytettävyystudkimuksen alalla on ollut useita yrityksiä vertailla käytettävyyssarviointimenetelmiä perustuen joihinkin yhteisiin kriteereihin [15][27]. Kuitenkaan ei olla pystytty muodostamaan mitään tapaa, jolla saataisiin luotettava arvio käytettävyyssarviointimenetelmien hyvyydestä. Tästä syystä tässä työssä käyttäjätutkimuksen menetelmiä vertaillaan ainoastaan niiden puutteiden ja etujen perusteella. Myös käytettävyyssalan ammattilaiset ja tutkijat [21] ovat sillä kannalla, että jokaisen käytettävyyssarviointimenetelmän ominaisuudet määrittävät menetelmän tehokkuuden löytää

tietynyyllisiä käytettävyysoongelmia. Näin ollen eri arviointimenetelmiä tulisi yhdistellä ottaen huomioon eri menetelmien puutteet ja vahvuudet. [21]

3.4 Heuristisen arvion ja näkökulmaan perustuvan arvion vertailu

Heuristisessa arvioissa verrataan palvelun jotain tiettyä ominaisuutta ennalta sovittuun heuristiikkaan. Tässä lähestymistavassa keskitytään vain itse ominaisuuksiin, eikä oteta huomioon käyttök kontekstia ja näin todenmukaisen arvion tekemisestä tulee hyvin vaikeaa [21]. Näkökulmaan perustuvassa arvioissa [26] sen sijaan verrataan tuotetta eri käyttäjärühmän tarpeisiin ja näin saadaan määritettyä myös käyttök konteksti. Jos vertaillaan näitä menetelmiä vain sen perusteella, mitä ne parhaimmillaan voivat tarjota, niin periaatteessa näkökulmaan perustuva arvio korvaa heuristisen arvion.

3.5 Heuristisen arvion, käytettävyydestä ja automaattisen käyttäjätiedon keräyksen vertailu

Käytettävyydestä tarjoo luotettavia arvioita, koska siinä on mukana otos oikean käyttäjärühmän käyttäjistä. Toisin kuin heuristisessa arvioissa ja muissa asiantuntija-arviointimenetelmissä käytettävyydestä saadaan tarkka kuva käyttäjän ja tuotteen välisistä interaktioista. Toisaalta on hyvin vaikeaa valita otosta käyttäjistä, joka kattaa koko käyttäjärühmän. Vääränlainen otos johtaa väärin päätelmiin käyttäjän tarpeista ja mieltymyksistä. On myös hyvin vaikeaa rajatussa ajassa kouluttaa käyttäjiä hallitsemaan tuotteen kehittyneimmät ominaisuudet; huonosti koulutetut käyttäjät voivat tehdä pinnallisia päätelmiä tuotteesta huomioidessaan ainoastaan tuotteen perusominaisuudet. Koska aika käytettävyydestä on yleensä rajallinen on vaikeaa jäljitellä hyvin realistisia käyttötilanteita, jossa loppukäyttäjä tuotetta mahdollisesti käyttäisi, sekä ottaa huomioon normaalissa käyttötilanteessa vaikuttavat motivaationaaliset tekijät, sekä oikeat päämäärät. Kun ei pystytä luomaan riittävän aitoa käyttöympäristöä tai käyttötilannetta eivät tuloksetkaan voi olla kovinkaan realistisia. Käyttäjien tarkkailu antaa myös vähän tietoa ongelmien syistä, koska siinä keskitytään vain siihen, mikä on tarkkailijoiden nähtävissä. [37]

Heuristinen arvio taas mahdollistaa tarkkailijoiden havainnoimien ongelmien taustalla olevien syiden tunnistamisen. Myös Doubleday et al. työ [28] tukee väitettä, että käytettävyydestä nähdään vain ongelman oireet, ei syytä. Sitä vastoin heuristisella arviolla löydetään ongelmien syyt mutta ei pystytä sanomaan kovinkaan paljo ongelmien oireista.

Toisaalta heuristisessa arvioissa keskitytään kerrallaan vain tiettyyn tuotteen ominaisuuteen, joka rikkoo käytettävyyssääntöjä, sekä syihin miksi näin tapahtuu. Voidaan saada myös selville miten näiden sääntöjen rikkomien tapahtuu, sillä tiedetään mikä tuotteen ominaisuus on sen aiheuttanut. Toisaalta heuristisen arvion heikkous on se, että arviot ovat todella subjektiivisia eri arvioijat eivät välttämättä tuota vertailukelpoisia tuloksia. On myös mahdollista, että käytettävyysoongelmia jää kokonaan löytämättä tai löydetään ongelmia, jotka eivät juuri vaikuta loppukäyttäjään. [37]

Brooksin [29] mukaan heuristinen arvio ei pysty ottamaan huomioon tuotteen kokonaisvaltaisempaa käsitettä eli hyväksyttävyyttä. Hyväksyttävyys koostuu hyödyllisyydestä,

käytettävyydestä, käytännön hyväksyttävyydestä ja sosiaalisesta hyväksyttävyydestä. Heuristinen arvio ei pysty myöskään ottaamaan huomioon käyttäjien toimien takana olevia ajatusketjuja tai käyttäjien mieltymyksiä. Tästä syystä ei pystytä sanomaan myöskään paljoa käyttäjien tarpeista [30]. Toisaalta käytettävyydestä ei pysty kertomaan pystyykö tuotteella pelkästään suorittamaan tehtävän vai tuottaako se käyttäjälle myös tyydytystä. Tällainen tieto on kuitenkin hyvin tärkeä tuotteen markkinoinnin kannalta. Näin ollen olisi hyvä myös etsiä tuotteesta hyviä puolia käytettävyyso Ongelmien lisäksi. Tästä syystä myös Molichin ja kumppaneiden CUE-4 tutkimuksessa [11] käytettävyysslaboratorioiden tuli kertoa myös positiivisista löydöistä .

On myös huomattava, että heuristinen arvio ei tarjoa mahdollisia ratkaisuvaihtoehtoja käytettävyyso Ongelmiin. Eri arvioijien tulokset on hyvin vaikea yhdistää, koska eri arvioijat näkevät ongelmat eri tavalla, joten on yleistä, että täysin sama ongelma on vain kuvattu eri tavalla [31]. Sen lisäksi heuristisella arviolla ei tule välttämättä käytyä koko tuotetta tai palvelua, sillä siinä ei ole mitään mekanismia, joka varmistaisi, että koko tuote tai palvelu on käyty läpi. Arvioijat voivat keskittyä liikaa myös johonkin tiettyyn osa-alueeseen. [30]. Heuristinen arvio ei myöskään kerro välttämättä sitä onko ongelma sellainen, että se esiintyy oikeassa käyttötilanteessa [32].

Automaattinen tiedonkeräys taas pyrkii ratkaisemaan useimmat käytettävyyssarvioon liittyvät ongelmat. Automaattisella tiedonkeräyksellä voidaan vähentää tarvetta käytettävyydestä ja välttää kattavan otoksen ottamista käyttäjäryhmästä ja sen tuomia haasteita. Tällöin saadaan myös analysoitua huomattavasti suurempi määrä erilaisia käyttäjiä ja näin ollen löydetty käytettävyyso Ongelmat ovat varmemmin sellaisia, jotka koskevat koko käyttäjäryhmää. Automaattisen tiedonkeräyksen ongelmana on se, että kerätty tieto ei juuri kerro käyttäjien käyttäytymisestä, heidän motiiveistaan tai odotuksista. [21]

3.6 Kognitiivisen läpikäynnin ja heuristisen arvion vertailu

Heuristinen arviolla saadaan kattava, mutta pinnallinen käsitys käyttöliittymästä. Siinä keskitytään käyttöliittymän eri osiin erikseen ottamatta huomioon oikeaa käyttök kontekstia. Kognitiivinen läpikäynti ottaa nimenomaan huomioon myös oikean käyttök kontekstin [31]. Se keskittyy kognitiivisiin asioihin, kuten opittavuuteen analysoimalla käyttäjien toiminnan taustalla olevia ajatusketjuja (mental models). Etuna tällä menetelmällä on se, että sen avulla voidaan saada selville käyttäjien päämääriä ja olettamuksia [30]. Kognitiivisen läpikäynnin etuna on myös se, että se keskittyy nimenomaan niihin tehtäviin joita käyttäjä suorittaa. Tämä auttaa suunnittelijoita arvioimaan kuinka heidän suunnitelmansa ominaisuudet sopivat käyttäjän käyttötilanteeseen. Kognitiivinen läpikäynti keskittyy tehtävän suorittamista vaativien tehtäväsarjojen toteuttamiseen joten se tarjoaa suunnittelijalle tietoa siitä, missä järjestyksessä asiat olisi hyvä käydä läpi tehtävän suorituksessa [32]. Mahdollisina haittapuolina on vääränlaisten tehtävien valinta, jonka johdosta tulokset antavat vääränlaisen kuvan tuotteen käytöstä.

3.7 Laboratoriotestauksen ja kenttätestauksen vertailu

Käytettävyydestä tarjoo hyvän tavan tunnistaa käytettävyyso Ongelmia yksityiskohtaisesti. Tästä syystä on myös paljon helpompaa rakentaa parannusehdotuksia ongelmien pohjalta. Käytettävyydestä tarjoo myös selkeää todistusaineistoa ongelman olemassaolosta. Toisaalta

varsinkin laboratoriossa suoritettujen kontrolloidun käytettävyydestä eivät välttämättä ota huomioon käyttäjän oikeaa käyttöympäristöä ja käyttötilannetta. Näin voidaan löytää ongelmia, jotka eivät oikeasti ole ongelmia oikeassa käyttötilanteessa. On myös vaikea valita käyttäjiä, jotka parhaiten edustaisivat tiettyä käyttäjäryhmää. Riskinä on myös se, että tutkijat vetävät liian kattavia johtopäätöksiä perustuen testauksessa olevalta pieneltä käyttäjämäärältä saatujen tietojen perusteella. [32]

Toisin kuin laboratoriotestauksessa, kenttätestauksessa pystytään ottamaan huomioon käyttäjän oikea käyttöympäristö ja käyttötilanne. Näin käyttäjä voi paremmin keskittyä tehtäväänsä unohtaen, että kyseessä on koetilanne ja näin koetilanteesta johtuva vääristävä vaikutus pienenee. Tämä voi auttaa myös siinä, että löydettyt ongelmat ovat oikeasti sellaisia, mitkä voivat tulla esille kun käyttäjä käyttää tuotetta työssään [32]. Kenttätestauksessa saadaan myös selville käyttäjien mieltymyksiä, sekä tietoa käyttäjän tyytyväisyydestä tuotteeseen.

Kenttätestauksen huonona puolena on se, että siinä vaaditaan huomattavan iso määrä käyttäjiä, jotta tuloksia voidaan pitää merkittävänä. Noin 30 käyttäjää pidetään minimimääränä luotettavalle tulokselle. Kenttätestauksella pystytään löytämään myös vähemmän ongelmia kuin muilla menetelmillä [30].

Kenttätestauksessa on myös mahdollista kerätä huomattavasti enemmän dataa pidemmällä ajanjaksolla kun laboratoriotestauksessa. Näin saadaan parempi kuva käyttäjän käyttäytymisestä kun päivän kuluessa opitaan kuinka hän normaalisti käyttää arvioitavaa tuotetta tai palvelua [32].

3.8 Äänen ajattelu -menetelmän ja kognitiivisen läpikäynnin vertailu

Tutkija Joe Dumas toteaa, että ääneen ajattelu -menetelmällä löydetään eniten käytettävyysongelmia [33]. Syy tälle voisi olla se, että se tarjoaa todella paljon dataa pienelläkin käyttäjämäärällä [30]. Ääneen ajattelu -menetelmällä voidaan saada tietoa siitä, mitä asioita on käyttäjän toimien taustalla. Kognitiivinen läpikäynti myös pyrkii vastaamaan samaan kysymykseen, mutta ääneen ajattelu -menetelmällä saadaan paljon syvällisempää tietoa itse ongelmista ja käyttökontekstista. Ääneen ajattelu -menetelmällä saadaan tarkka approksimaatio siitä, miten yksittäiset käyttäjät käyttävät tuotetta yleensä. Tämä tieto sisältää käyttäjien kommentteja, jossa käyttäjät kertovat elävästi ja tarkasti tuotteeseen liittyvistä kysymyksistä, sekä käyttäjien mieltymyksiä. Haittapuolena on se, että se ei tue ihmisiä, joiden oppiminen poikkeaa analyyttisestä lähestymistavasta. Epäanalyyttinen oppija voi tuntea olonsa epäluonnolliseksi ja se vaikuttaa hänen suoritukseensa käytettyystestissä [30].

3.9 Yhteenveto vertailusta

Taulukossa 1. pyritään kasaamaan edellä mainittuja käytettävyystudkimusmenetelmien eroja. Huomionarvoista on, että ääneen ajattelu -menetelmään vaikuttaa käytettävyydestä testauksen ominaisuudet, sillä se suoritetaan aina käytettävyydestä testauksen yhteydessä. Eroavaisuuksia silti on, sillä ääneen ajattelu -menetelmällä saatu tieto on kvalitatiivista ja käytettävyydestä testauksella kerätään kvantitatiivista dataa, jolloin saadaan huomattavasti helpommin tilastollisesti merkittävää dataa. Taulukossa perinpohjaisuudella tarkoitetaan sitä, että miten syvällisesti menetelmä käy arvioinnin kohteena olevat ongelma-alueet. Laaja-alaisuudella taas tarkoitetaan

sitä kuinka monia eri ongelma-alueita se käy läpi. Tieto ongelmasta osiosta kertoo, että ongelmasta voidaan saada selville sen oireet ja/tai syyt.

Käytettävyydestit ovat hyvin läpikotaisia, sillä niissä keskitytään tarkkailemaan oikeita käyttäjiä oikeissa tehtävissä. Vaikka niissä kerätään usein kvantitatiivista dataa on usein liian kallista tehdä niistä tilastollisesti kattavia. Automaattiset käyttäjädatankeräys on taas hyvin pintapuolista, koska siinä voidaan määrätä vain hyvin rajallinen määrä suureita mitattavaksi. Toisaalta käyttämällä webin käyttö -analyysiä saadaan tarkempaa tietoa käyttäjien toimista. Näkökulmaan perustuva arvio taas on vähemmän kattava kuin heuristinen arvio, sillä siinä kapeammin tiettyihin ongelma-alueisiin eri näkökulmista.

Taulukko 1. menetelmien vertailua

	Perinpohjaisuus	Laaja-alaisuus	Tieto ongelmista	Ympäristö (ottaako huomioon)	Konteksti (ottaako huomioon)	Tilastollinen merkitys
UT	Hyvin läpikotainen	suppea	oireet	vähän	kyllä	harvoin
UT2	Hyvin läpikotainen	suppea	oireet	kyllä	kyllä	harvoin
TA	Läpikotainen	suppea	syyt ja oireet	vähän	kyllä	ei ole
TA2	Läpikotainen	suppea	syyt ja oireet	kyllä	kyllä	ei ole
HE	pintapuolinen	kattava	syyt pintapuolisesti	ei	ei	ei ole
PBE	Läpikotainen	vähemmän kattava	syyt	ei	kyllä	ei ole
DL	Hyvin pintapuolinen	suppea	vähäinen	ei	ei	on
DL2	pintapuolinen	suppea	hieman tietoa oireista	ei	ei	on
CW	Läpikotainen	suppea	syyt	ei	kyllä	ei ole

[UT = käytettävyydestaus laboratoriossa, UT2 = käytettävyydestaus kentällä, TA = ääneen ajattelu -menetelmä (think aloud) laboratoriossa, TA2 = ääneen ajattelu -menetelmä kentällä, HE = heuristinen arvio, PBE = näkökulmaan perustuva arvio (perspective-based evaluation), DL = automaattinen käyttäjädatan keräys (data logging), DL2 = automaattinen datankeräys + webin käyttö -analyysi, CW = kognitiivinen läpikäynti (Cognitive Walkthrough)].

3.10 Kuinka käytettävyydestutkimuksen menetelmiä tulisi käyttää yhdessä, jotta ne paikkaisivat toistensa puutteita

Kuten Matera et al. työssä [21] todettiin, niin paras tulos käytettävyydenarviossa saadaan yhdistelemällä eri menetelmiä ottaen huomioon niiden hyvät ja huonot puolet. Tässä työssä todettiin myös, että eri käytettävyydenarviointimenetelmillä löydetään erilaisia ongelmia.

On selvää, että tuotekehitysprosessissa aikaa vievämpiä menetelmiä, kuten käytettävyydestausta ei voida käyttää montaa kertaa ajanpuutteen ja kustannusten takia [21]. Tästä syystä on välttämätöntä käyttää myös menetelmiä, joita voi käyttää iteratiivisesti tuotekehitysprosessin aikana [17]. Vaatimuksena näille iteratiivisille menetelmille on, että niiden käytön tulee olla mahdollisimman nopeaa kuitenkin siten, että ne tarjoavat riittävän luotettavia tuloksia. Heuristinen arvio on riittävän kevyt menetelmä tähän käyttötarkoitukseen. Sen lisäksi työ [36] yrittää löytää vastauksia siihen miten yksinkertainen käytettävyydenarviointimenetelmä voi olla, jotta siitä mitään hyötyä. Työssä esitetään idea, että pyritään löytämään kerralla vain joitakin vakavia

käytettävyysoongelmia. Koska työssä käytettävää menetelmää on todella nopeaa käyttää voidaan tuotetta arvioida iteratiivisesti löytäen aina uusia ongelmia eri arviointikerroilla. Tosin hyvin yksinkertaisissa menetelmissä on myös ongelmansa. Ongelma on siinä, että kun käytetään todella pientä näytekokoja ei voida tietää onko ongelman vaikutus tuotteeseen todella riittävän laaja, jotta se tulisi paikata. Näin ollen on tärkeää ottaa huomioon, mitä oikeastaan voidaan taata käyttämällä työssä esitettyä menetelmää. On sanomattakin selvää, että tällaisten äärimmillen vietyjen menetelmien käyttö vaatii laajaa asiantuntemusta käytettävyytutkimuksesta ja ymmärrystä siitä, mihin näillä menetelmillä oikeastaan pystyy. Tämän artikkelin pohjalta voitaisiin määrittää optimaalinen menetelmä nimenomaan iteratiivisen tarkastelun tueksi.

Kun nopeilla iteratiivisilla menetelmillä, kuten heuristisella arviolla havaitaan ongelmia käytettävyydessä, jotka koetaan riittävän vakaviksi voidaan ottaa käyttöön aikaa vievämpi ja tehokkaampi menetelmä kuten käytettävyystestaus. Myös työssä [37] ehdotetaan, että heuristisella arviolla voisi ensin saada pintapuolisemman, mutta laajemman kuvan tuotteen ongelmista, jonka perusteella voisi päättää, mitä asioita tulee testata käytettävyydestä. Tarkistusmenetelmät, jotka tarjoavat nopeasti käsityksen mahdollisista käytettävyysongelmista auttavat luomaan kokonaiskuvaa siitä, mihin läpikotaisemmassa käytettävyyttestauksessa tulisi keskittyä ja mitä palvelun osa-alueita tulisi ottaa testaukseen mukaan. Kantner, L et al.[22], ehdottavat juuri tällaista ratkaisua: käytetään nopeaa heuristista arviota tuottamaan käsitys mahdollisista käytettävyysongelmista ja sen pohjalta päätetään mihin palvelun osa-alueisiin keskitytään käyttäjättestauksessa.

Toisaalta käytettävyyttestaus keskittyy yksityiskohtaisesti vain valittujen tehtävien avulla läpikäytyihin asioihin, joten sen tukena hyvä käyttää menetelmiä, jotka antavat laajemman, mutta vähemmän yksityiskohtaisen kuvan tuotteen käytettävyysongelmista. Tästä syystä heuristinen arvio olisi sopiva myös tähän tarkoitukseen [32]. Heuristinen arvio sopisi myös hyvin käytettäväksi kognitiivisen läpikäynnin kanssa, sillä myös kognitiivinen läpikäynti keskittyy vain valituista tehtävistä nousseisiin asioihin [30].

Käytettävyyttestauksen tukena on hyvä käyttää myös haastattelua varsinkin silloin jos käytettävyyttestaus suoritetaan kenttätestauksena. Näin saadaan arvokasta tietoa liittyen siihen kuinka käyttöympäristö vaikuttaa käyttäjän toimiin hänen käyttäessään testavaa tuotetta. Haastattelulla saadaan myös selville jos jokin ominaisuus ei miellyttänyt käyttäjää, vaikka se ei hänen työskentelyn tulokseen vaikuttanutkaan. Saadaan myös selville mitä ominaisuuksia käyttäjä suosii käyttäessään tuotetta ja miten tyytyväinen käyttäjä on yleensä tuotteen toimintaan.

Kun käytettävyyttestä suoritetaan kenttätestauksena saadaan yksityiskohtaista tietoa, siitä miten käyttäjä käyttää tuotetta oikeassa ympäristössä niin tällöin ne tuotteen osa-alueet, joihin keskitytään, rajoittuvat pelkästään valittuihin tehtäviin. Tästä syystä kun sen tukena käytetään laajemman, mutta pinnallisemman kuvan antavaa menetelmää kuten heuristista arviota, saadaan huomattavasti kattavampi kuva tuotteen kokonaiskäytettävyydestä [32]. Sen lisäksi, kuten menetelmien vertailussa mainittiin, käytettävyyttestaus ei juuri anna tietoa ongelmien syistä, koska siinä keskitytään vain siihen, mikä on tarkkailijoiden nähtävissä [21]. Näin ollen sen tueksi on hyvä ottaa menetelmä, joka mahdollistaa ongelmien taustalla olevien syiden tunnistamisen, kuten heuristinen arvio [21].

Koska käytettävyyttestaus ja heuristinen arvio eivät juuri kerro mitään käyttäjän toimien taustalla olevista ajatusmalleista, niin samanaikaisesti on mahdollista suorittaa ääneen ajattelu -menetelmällä tapahtuva arvio, joka nimenomaan keskittyy siihen, miksi käyttäjät toimivat niin kuin toimivat [30].

4 Johtopäätökset

4.1 Mitä voidaan tehdä, jotta käytettävyytutkimusmenetelmillä saataisiin parempia tuloksia

Kuten tutkimusaineisto-osiossa kävi ilmi, käytettävyytutkimuksen menetelmillä löydetään todella erilaisia käytettävyysoongelmia. Sen lisäksi käytettävyysoongelmia löydetään määrällisesti todella paljon enemmän kuin niihin voidaan puuttua. Näiden ongelmien pohjalta tässä työssä pyritään esittämään parannuksia nykyisiin käytettävyytutkimusmenetelmiin, jotta niillä saadut tulokset olisivat yksinkertaisesti parempia.

4.2 Käyttäjien ja käytettävyystestaaajien määrän merkitys

Nielsen väitti vuonna 2000 ilmestyneessä artikkelissaan [38], että käytettävyydesti, jossa on mukana vain 5 käyttäjää paljastaa 85% palvelun kaikista käytettävyysongelmista. Kaikilla käytettävyysongelmilla tarkoitamme tässä kaikkia löydettyjä käytettävyysoongelmia yhteensä. Ei voida oikeasti missään vaiheessa sanoa, että kaikki ongelmat olisivat löytyneet. Molichin ja kumppaneiden CUE-2 [10] ja CUE-4 [11] kuitenkin osoittavat, että 5 käyttäjällä löydettyjen käytettävyysongelmien määrä on todella paljon pienempi. CUE-2 tutkimuksessa yksi tiimi löysi parhaimmillaan 75 % kaikista 310:stä ongelmasta ja CUE-4 tutkimuksessa vain 67 % kaikista 237:stä ongelmasta. Eli nykyisillä menetelmillä pystytään löytämään vain murto-osa kaikista käytettävyysongelmista eikä läheskään niin paljon kuin Nielsen väittää.

Kun CUE-2 -tutkimuksessa vertailtiin käytettävyysteisteissä olevien käyttäjien määrään ei huomattu sen vaikuttavan juurikaan löydettyjen ongelmien määrään eikä vakavuuteen. Tiimi joka käytti 50:tä käyttäjää löysi vähiten ongelmia kun taas tiimi, joka käytti vähiten käyttäjiä löysi toiseksi vähiten ongelmia. CUE-2 tutkimuksessa todetaan loppujen lopuksi, että käytettävyydestillä, jossa on viisi käyttäjää löydetään riittävä määrä ongelmia ja käyttäjämäärän kasvattaminen viidestä ylöspäin ei juuri tuo lisäetua.

Työssä [39] tutkittiin kokeellisesti, miten käyttäjien määrän lisääminen vaikuttaa löydettyjen ongelma-alueiden prosenttiosuuteen. Siinä huomattiin, että kasvattamalla käyttäjämäärää yli viiden vähennetään löydettyjen ongelma-alueiden keskihajontaa, eli löydetään suuremmalla varmuudella vastaava määrä ongelmia. Tässä tutkimuksessa ei kuitenkaan huomattu, että käyttäjämäärän lisääminen olisi lisännyt löydettyjen ongelmien määrää aivan kuten Molichin ja kumppaneiden CUE-4 tutkimuksessa [11]. Huomattiin tosin, että erilaisten käytettävyystestien tehtävien määrällä on selkeä vaikutus löydettyjen ongelmien määrään. Myös CUE-4 -tutkimus [11] tukee tätä väitettä. Nämä testit toki perustuivat kattaviin käytettävyysteisteihin (60 osallistujaa), mutta eri käyttäjämäärän vaikutusta simuloitiin Matlab -ohjelmistolla käyttäjämäärien liikuessa välillä 5-60 tehden 100 simuloitua käytettävyyssajoa jokaiselle käyttäjämäärälle, joten testin luotettavuudesta voidaan olla montaa mieltä. Voidaan kuitenkin sanoa, että käyttäjämäärän lisäämistä ei voida pitää parhaana ratkaisuna kun halutaan löytää enemmän käytettävyysoongelmia, sillä käyttäjämäärän lisääminen ainoastaan parantaa

todennäköisyyttä löytää enemmän ongelmia, eikä takaa sitä.

Cocktonin ja kumppaneiden artikkelissa käsitellyn tutkimuksen mukaan [40] käytettävyydestaajien lisäämisellä on vaikutusta löydettävien käytettävyysongelmien määrään. Myös Molich ja kumppanit ovat samoilla linjoilla: kun käytettävyyssarvioijien määrää lisätään, niin löydetään yhä enemmän erilaisia käytettävyyso ongelmia. Cocktonin ja kumppaneiden tutkimuksessa käytettävyydestien vertailun apuna käytettiin käsitteitä tarkkuus (1) ja validiteetti (2).

$$\text{Tarkkuus} = \frac{\text{arvioijien löytämät käytettävyyso ngelmat}}{\text{kaikki löydetyt käytettävyyso ngelmat}} \quad (1)$$

$$\text{Validiteetti} = \frac{\text{oikeat käytettävyyso ngelmat}}{\text{väärät hälytykset}} \quad (2),$$

väärät hälytykset ovat asioita, jotka on virheellisesti luokiteltu käytettävyyso ngelmiksi kuitenkin olematta niitä.

Käytettävyydestaajien määrän kasvaessa arvioinnin tarkkuus (1) ja läpikotaisuus kasvaa asymptoottisesti, mutta toisaalta kun käytettävyyssarvioijien määrä kasvaa, niin heidän on yhä vaikeampi muodostaa yhteisymmärrys eri ongelmien vakavuustasoista ja näin ollen kokonaisarvion validiteetti (2) heikkenee ja sen seurauksena arvion teho heikkenee. Tärkeä huomio tässä tutkimuksessa on, että kun käytettävyydestaajien määrää lisätään 7:stä niin käytettävyyssarvion teho ei enää nouse vaan päin vastoin; se heikkenee käytettävyyssarvioijien määrän ylittäessä kahdeksan henkilön rajan. Näin ollen seitsemän käytettävyyssarvioijaa olisi hyvä määrä tehokkaimman arvion saavuttamiseksi.

Käyttäjien ja käytettävyydestaajien määrään vaikuttaa tosin hyvin paljon arvioitava palvelu. Molich ja kumppanit [11] huomauttavat, että jos kyseessä on esimerkiksi sairaalan laitteisto voi olla hyvin tärkeää löytää vielä lisää käytettävyyso ngelmia. Tämä esimerkiksi siitä syystä, että valideetin laskiessa joudutaan käyttämään yhä enemmän työtunteja siihen, että eri käytettävyyssarvioijat saavuttavat yhteisymmärryksen löydetyistä ongelmista.

Toisaalta tutkimus [41] osoittaa, että käytettävyydesteissä oikeiden tehtävien valinta on paljon tärkeämpää kuin oikean käyttäjämäärään määrittäminen. Näin ollen huomion pitäisi enemmän keskittyä siihen, että käytettävyydestien tehtävien valinnassa keskitytään ainoastaan tärkeimpiin asioihin, sekä suunnitellaan käytettävät tehtävät huolella. Myös tutkimus [42] toteaa saman asian. Sen lisäksi siinä todetaan, että erilaiset tehtävät käytettävyydesteissä tuottavat erilaisia tuloksia. Tutkimuksen mukaan, jos tehtävä on tarkasti suunniteltu ja hyvin rakenteellinen se paljastaa enemmän käytettävyyso ngelmia, mutta tällöin löydetyt ongelmat ovat pääasiallisesti vähäpätöisiä ja pinnallisia. Toisaalta vapaammin toteutetut ja suunnitellut tehtävät taas löytävät vähemmän vähäpätöisiä ongelmia ja huomattavasti enemmän todella vakavia käytettävyyso ngelmia. Eli sellaisia ongelmia, jotka voivat johtaa siihen, että käyttäjä lopettaa tuotteen käytön kokonaan.

4.3 Tarvitaan ongelmien löytämiseen lisää järjestelmällisyyttä

Molichin ja kumppaneiden CUE-2 [10] ja CUE-4 [11] tutkimuksissa eri käytettävyyssiimit löysivät todella erilaisia käytettävyyso ngelmia, eikä arviointiin käytetyllä työmäärällä ei tuntunut olevan vaikutusta löydettyjen ongelmien määrään, oli menetelmä mikä tahansa. Näiden

tulosten valossa voidaan todeta, että käytettävyyden arviointimenetelmät tuottavat hyvin sekalaisia ja epä johdonmukaisia tuloksia, mikä vaikeuttaa lopullisen arvion tekemistä huomattavasti. Jotain tulisi tehdä, jotta käytettävyyden arviointimenetelmillä pystyttäisiin arvioimaan palvelun käytettävyyttä huomattavasti järjestelmällisemmin ja syvällisemmin kuitenkin lisäämättä työmäärää merkittävästi. CUE-2 [10] ja CUE-4 [11] tutkimuksissa käytettävyydelaboratorioiden käyttämät menetelmät voidaan laskea kuuluvan karsittuihin käytettävyyden arviointimenetelmiin. Nämä menetelmät ovat muutenkin vallitsevassa asemassa tämän päivän käytettävyydestauksessa, kuten CUE-2 [10] ja CUE-4 [11] tutkimukset osoittavat. Karsittujen käytettävyyden arviointimenetelmien käyttö voi olla juuri yksi syy niillä löydettyjen käytettävyysongelmien suuren eroavaisuuteen. Tätä tukee myös Cocktonin ja kumppaneiden [40] väite, että mikään karsituista käytettävyyden arviointimenetelmistä ei käy systemaattisesti läpi kaikki mahdollisia osa-alueita, joilla käytettävyyden ongelmia voisi ilmetä.

Käytettävyydsarvion subjektiivisuus on luultavasti yksi perussyy miksi CUE-4 [11] tutkimuksessa eri käytettävyydsarvioijat löysivät niin erilaisia ongelmia. Hertzumin ja Jacobsenin työ [17] ottaa myös esille, miten suuri vaikutus käytettävyydsarvioijalla itseasiassa on testin tulokseen. Sen lisäksi siinä ehdotetaan, kuinka eri käytettävyydsarvion tutkimusmenetelmissä arvioijan vaikutus tulisi ottaa huomioon.

Käytettävyydesteissaan yleistä, että arvioija valitsee itse tehtävät, joita käyttäjät tulevat suorittamaan tuotteella [43]. Näin arvioijalla on suuri vaikutus lopputulokseen, sillä tehtävien valinta on todella tärkeä osa käytettävyydestin onnistumista [40]. Varsinkin kvantitatiivissa mittauksissa tämä ongelma on hyvin ilmeinen. Arvioijan vaikutusta voidaan estää ottamalla käyttäjä mukaan tuotteella suoritettavien tehtävien suunnitteluun.

Automaation lisääminen on yksi varteenotettava vaihtoehto systemaattisuuden lisäämisessä käytettävyydsarvion tutkimukseen [44]. Toisaalta automatisoidut työkalut on perinteisesti tarkoitettu käsittelemään kvantitatiivista dataa, jonka jälkeen ollaan hyödynnetty tilastollisia analysointimenetelmiä. Ongelma on kuitenkin siinä, että harvoin käytettävyydestauksessa saadulla datalla on tilastollista merkitystä, sillä käytettävyydesteissa olevat käyttäjämäärät ovat liian pieniä. Tilastollisella merkityksellä tarkoitetaan käytännössä sitä todennäköisyyttä, että ei tehdä virhepäätelmiä. Vaikka käytettävyydestauksella saatu data ei ole lähellekään tilastollisesti merkitsevää, parantaa se kuitenkin aina todennäköisyyttä, että tehdään oikeita päätelmiä, jotka parantavat palvelun käytettävyyttä. [21].

Jos käytettävyydsarvion tutkimuksessa haluttaisiin tilastollisesti merkittävää dataa tulisi käyttää automatisoituja tiedon keräysmenetelmiä. Automaattinen tiedon kerääminen webissä on yksi varteenotettava vaihtoehto tähän (katso käytetyt menetelmät osiossa 3.). Näin laaja datamäärä auttaa myös hahmottamaan oikeamman kokonaiskuvan käyttäjistä. Kun pystytään analysoimaan lähes kaikkia palvelua koskevan käyttäjäryhmän jäseniä saadaan myös enemmän tiettyä käyttäjäryhmää koskevia ongelmia. Toisaalta automaattisten datankeräämismenetelmien tarjoama informaatio ei kerro juuri mitään asioiden taustoista tai merkityksestä, kuten käyttäjien päämääristä palvelua käytettäessä ja heidän odotuksistaan, jotka ovat hyvin keskeisiä kysymyksiä käytettävyydsarviossa [21].

Vaikka menetelmä nimeltä korttien lajittelu (card sorting) on menetelmänä hyvin epätieteellinen ja vaikeasti tulkittava, se kertoo siitä miten käyttäjä hahmottaa laajan sivun kokonaisuutena. Muut käytettävyydsarvion tutkimuksen menetelmät eivät juuri valota tätä asiaa. Tässä menetelmässä käyttäjille annetaan kasa kortteja johon on kirjoitettu sivustolla sijaitsevia otsikoita ja käyttäjä pyydetään järjestämään nämä mielestään loogisiin kokonaisuuksiin. Nykyiset www-sivut ovat hyvin laajoja, näin ollen ei riitä, että käyttäjä pystyy suoriutumaan joistakin sivustoon liittyvistä tehtävistä. Tämä ei kerro siitä, pystyykö käyttäjä hahmottamaan sivuston kokonaiskuvan, sivun

eri osien väliset suhteet, menettelytavat joilla suoriudutaan monimutkaisista tehtävistä hyödyntäen tietoa sivun eri osien välisistä suhteista ja kokonaiskuvasta. Käytettävyydestit keskittyvät yleensä kerrallaan vain palvelun yhteen osa-alueeseen. Näin kun käyttäjä joutuu hyödyntämään tietoa sivun eri osien välisistä suhteista ominaisuudet, jotka yksittäisissä tehtävissä toimivat eivät toimikaan enää halutulla tavalla. Tällöin voi käydä niin, että käytössä olevat käytettävyydestit eivät käy riittävän laajasti läpi sivun eri osa-alueita voi jäädä huomaamatta todella merkittäviä käytettävyyso ongelmia. [47]

Koska nykyiset verkkosivustot ovat laajoja, tarvitaan automatisoituja menetelmiä, joilla pystytään selvittämään millä tietyillä alisivuilla käyttäjät kokevat eniten käytettävyyso ongelmia. Apua tähän tuo työssä [52] esitetty käyttäjien verkkosivujen selaamista tarkkaileva työkalu. Tämän työkalun avulla voidaan helposti visualisoida ne polut joita pitkin eri käyttäjät kulkevat selaaillessaan sivuston sisältöä. Tämän työkalun avulla on myös mahdollista tutkia miksi jotkut ihmiset lähtevät sivulta ennen aikojaan. Näiden tietojen avulla voidaan pyrkiä selvittämään, millä alisivulla olisi mahdollisesti eniten käytettävyyso ongelmia ja keskittyä pelkästään sen käytettävyyso ongelmien korjaamiseen.

Jos Internetistä oikeasti haluttaisiin kerätä merkityksellistä dataa automaattisesti, niin koko Internetin luonteen tulisi muuttua. Semanttinen web voisi tarjota tähän ratkaisun, sillä sen peruseriaatteisiin kuuluu se, ettei Internetissä liiku pelkkää raakaa dataa vaan sille on erikseen koodattu merkityksiä [46].

Yksi syy siihen, miksi käytettävyydestit eivät ole tehokkaita, on kommunikaation puute käytettävyyso asiantuntijoiden ja suunnittelutiimien välillä. Myös Molich ja kumppanit toteavat työssään [11], että osa käytettävyyso ongelmista on yleensä suunnittelutiimin tiedossa, mutta niitä ei kerrota johtohenkilöstölle ja näin tieto ei leviä heidän kauttaan käytettävyyden arvioijille. Työssä [51] todetaan, että kun tuotekehitystiimin jäsenet eivät ole mukana suunnittelussa ja arviointiprosessissa tai käytettävyyso arvioijilla ei ole mahdollisuutta työskennellä suunnittelijoiden kanssa, tuotteeseen tehtyjen parannusten määrä vähenee huomattavasti. Toisaalta, jos tuotekehitystiimi on mukana luottamus ja kunnioitus tuotekehitystiimin ja käytettävyyso asiantuntijoiden välillä kasvaa. Näin ollen tuotekehitystiimiläisten kynnys ottaa käytettävyyso asiantuntijoiden kommentit huomioon suunnittelussa alenee huomattavasti. Tällöin tuotekehitystiimi pääsee myös heti näkemään käytettävyyso ongelmat ja keskustelemaan niistä käytettävyyso asiantuntijoiden kanssa ja tuotteeseen kaavailut parannusehdotukset voidaan toteuttaa välittömästi. Näin myös käytettävyyso asiantuntijat pääsevät nopeammin arvioimaan muuttunutta palvelua. Kun tuotekehitystiimin ja käytettävyyso asiantuntijoiden yhteistyö on riittävän tiivistä, tarve kirjoittaa aikaa vieviä raportteja vähenee ja voidaan hyödyntää nopeampia menettelytapoja. Ohjeita tämän toteuttamiseen löytyy useiden alan johtavien käytettävyyso tutkijoiden työstä [51].

Vaikka toisaalta, jos käytettävyyso arvio käy systemaattisesti läpi kaikki mahdolliset osa-alueet, joilla voi ilmetä käytettävyyso ongelmia, tullaan varmasti löytämään yhä enemmän ongelmia. Molich ja kumppanit [10][11] osoittivat, että jo nyt käytössä olevilla karsituilla käytettävyyden arviointimenetelmillä löydetään aivan liikaa ongelmia. Silti tämä on välttämätön paha, jos oikeasti halutaan saada laaja käsitys kaikista käytettävyyso ongelmista. Helpotusta tähän ongelmalliseen tilanteeseen toisi se, jos pystyttäisiin luokittelemaan ja karsimaan vähäpätöisiä käytettävyyden ongelmia erittäin tehokkaasti. Juuri tätä kysymystä käsitellään seuraavassa luvussa.

Kovinkaan kokonaisvaltaista automaatiota ei ole hyvä käyttää itse tutkimusmenetelmien toteutuksessa, sillä nykyisillä menetelmillä tällöin menetettäisiin todella paljon oleellista subjektiivista dataa. Automaatiota voidaan tosin hyödyntää siinä vaiheessa kun

käytettävyyssasiantuntijat ovat keränneet perinteisillä menetelmillä riittävän otoksen raakaa käytettävyyssdataa. Tämän jälkeen voidaan hyödyntää datan louhintatekniikoita mielenkiintoisten yhteyksien löytämiseen käytettävyyssongelmien välillä. Tästä lisää seuraavassa osiossa.

4.4 Tarvitaan tehokkaampia keinoja luokitella ja karsia vähäpätöisiä ongelmia

Perinteisesti käytettävyyssastauksessa on kerätty kvantitatiivista dataa, josta ollaan nostettu esille tärkeitä asioita tilastollisen analyysin avulla. Kuitenkin kvantitatiivisen datan analyysissä sivuutetaan tärkeitä asioita, kuten käyttökonteksti, käyttöympäristö ja käyttötarkoitus. Kun taas kerätään dataa kvalitatiivisilla käytettävyyss tutkimusmenetelmillä ja kerätään saatu data tietokantoihin (yksinkertaisimmillaan taulukkomuotoon) ongelmaksi tulee se, mikä on oleellista tässä massivisessa datamäärässä.

Hornbæk et al. [48] painottavat, että ongelmien luokittelu on todella tärkeää, koska sen avulla pystytään tutkimaan käytettävyyssarviointimenetelmien suhteellista tehokkuutta, sekä selvittämään tärkeitä ilmiöitä kuten arvioijan vaikutusta tulokseen. Kuten tutkimusaineisto-osiossa mainittiin Molich ja kumppanit havalsivat CUE-4 tutkimuksessaan [11], että riippumatta käytettävyyden arviointimenetelmästä käytettävyyssongelmia löydetään huomattavasti enemmän kuin niitä pystytään käsittelemään. Tämä on mielestäni tärkein käytännön syy luokittelun tärkeydelle. CUE-4 [11] osoittaa myös, että hyvien luokitteluperusteiden laatiminen on vaikeaa, jopa tämän hetken johtaville käytettävyyssasiantuntijatiimeille. Näissä tutkimuksissa tiimejä pyydettiin luokittelemaan löytämänsä käytettävyyssongelmat vakavuusjärjestykseen. Parhaimmillaan yksi tiimi löysi 35 ongelmaa 50 vakavimman ongelman joukosta. Tiimejä pyydettiin myös määrittämään 5 vakavinta käytettävyyssongelmaa. Mukana oli kaiken kaikkiaan 17 tiimiä ja niistä 7 tiimin vakavimman ongelman listasta yksikään ei löytynyt muiden 10 tiimin listasta. Näin ollen on selkeä tarve kehittää parempia käytettävyyssongelmien luokitteluperusteita.

Kun käytetään tehokkaita ja standardoituja luokittelumenetelmiä niin eri arvioijien on myös helpompi muodostaa yhteisymmärrys ongelmien luokittelusta. Tällöin kun lisätään käytettävyyssarvioiden määrää, löydetään enemmän ongelmia, mutta paremman luokittelun avulla validiteetti (2) karsii paljon vähemmän.

Hornbæk et al. [48] esittelevät yhden tavan luokitella käytettävyyssongelmia. Työssä otetaan esille neljä eri luokitteluperustetta.

- Ensimmäinen tapa on luokitella sellaiset ongelmat, jolla on samanlainen vaikutus tuotteeseen kuuluvaksi samaan ryhmään.
- Toinen tapa on luokitella ongelmia niiden arvioidun vakavuuden mukaan, eli luokitella ongelmat vakavimmasta vähiten vakavaan.
- Kolmas peruste on Lavery et al. kehittämä malli, joka perustuu olettamukselle, että jokainen käytettävyyssongelma koostuu neljästä eri komponentista. Nämä komponentit ovat:
 - Syy, esimerkiksi ongelman aiheuttanut suunnitteluvirhe.
 - Virheellinen toiminta, kuten ongelmasta johtuen käyttäjä tulkitsee käyttöjärjestelmän antaman palautteen väärin.

- Toiminnan tulos, kuten ongelman takia epäonnistunut tehtävä.
- Suunnitelmaan tuleva muutos, eli muutetaan jotain käyttöliittymän ominaisuutta ongelman poistamiseksi.

Ideana on järjestellä käytettävyysongelmat vakavuusjärjestykseen sen mukaan kuinka kuinka monta eri komponenttia koskee yhtä käytettävyysongelmaa.

- Neljäntenä vaihtoehtona käytettävyysongelmien luokittelulle on käyttäjien käytön arviointia auttava valmis malli (user action framework). Se on tietokanta, joka sisältää käytettävyyteen liittyviä käsityksiä ja kysymyksiä. Tietokanta koostuu hierarkkisesta puurakenteesta, jonka eri tasot muistuttavat hieman Normanin seitsemän vaiheen mallia (katso näkökulmaan perustuva arvio osiossa 3). Tämä malli koostuu kategorioista kuten:
 - Suunnittelukysymykset.
 - Käyttöliittymässä olevien käsitteiden tulkitseminen.
 - Fyysiseen toimintaan liittyvät kysymykset.
 - Käyttäjän toiminnan tuloksiin liittyvät kysymykset.
 - Systemin toiminnallisuus.
 - Systemin arviointi.
 - Ongelmat, jotka ovat riippumattomia prosessin iteraation vaiheista.

Ideana on järjestellä saadut käytettävyysongelmat tämän hierarkkisen käsitepuun oksille. Kun kaikki ongelmat on järjestelty tulee vielä miettiä vastaako se käsite, minkä alle ongelma on asetettu parhaiten sen ongelman luonnetta. Näin ollen tuloksena saadaan ongelmakategorioita, joiden otsikoina on tämän valmiin mallin käsitteet. Eli ongelmat, jotka laitetaan eri kategorioihin ovat eri ongelmia ja samaan kategoriaan laitettut ongelmat samoja ongelmia.

Tässä työssä vertailtiin näiden neljän eri luokittelutavan hyvyttä. Havaittiin samankaltaisia ilmiöitä, kuten Molichin ja kumppaneiden CUE-tutkimuksissa, eli eri luokittelumenetelmillä eri testaajat saivat hyvin erilaisia ongelmakokonaisuuksia, joten myös näillä luokitteluperusteilla esiintyy ongelmia liittyen menetelmien yksiselitteisyyteen.

- Kun arvioitiin ensimmäistä luokittelutapaa eli samankaltaisuuden mukaan yhteen luokittelua havaittiin, että tällä menetelmällä muodostuu eniten erillisiä ongelmaryhmiä. Havaittiin myös, että luokittelijoilla oli suuria vaikeuksia päättää kuuluuko ongelma johonkin kokonaisuuteen vai tuleeko muodostaa uusi kokonaisuus.
- Kun arvioitiin toista luokittelutapaa eli käytännön priorisointia saavutettiin paljon parempia tuloksia. Tämä menetelmä saavutti suurimman yhteisymmärryksen luokitteluperusteista eri luokittelijoiden välillä. Tällä menetelmällä käyttäjät perustivat havaintonsa omiin kokemuksiinsa, jonka olisi luullut tuottavan erimielisyyksiä luokittelijoiden kesken, mutta tällaista ei kuitenkaan havaittu.
- Kun arvioitiin kolmatta luokittelutapaa eli Lavery et al. mallia eri luokittelijat saavuttivat pienimmän yhteisymmärryksen ongelmien luokittelun suhteen vertailussa muihin tapoihin. Yksi syy tähän lienee ollut se, että kokemattomat luokittelijat eivät olleet kunnolla sisäistäneet tätä mallia lyhyen opetteluun jälkeen.
- Kun arvioitiin neljättä luokittelutapaa eli ”user action framework”:ia huomattiin, että se tuotti paljon eri ongelmaluokkia. Luokittelijat kokivat, että tällä menettelytavalla luokittelu oli haastavinta, mutta myös tässä kuten edellisessä tavassa syy lienee se, että

näin monimutkaista mallia ei pysty sisäistämään testiin varatussa ajassa. Toisaalta tutkimuksessa todettiin, että tämä menetelmä auttaa saamaan syvällisemmän ymmärryksen käytettävyyden ongelmista.

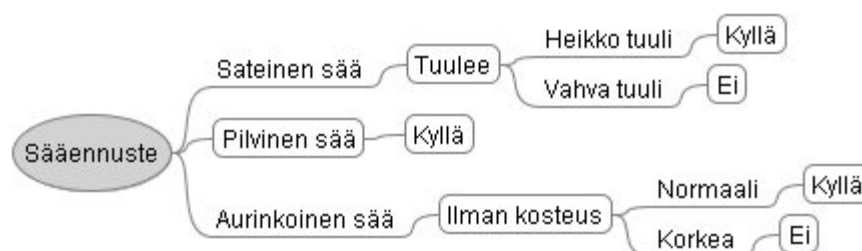
Tässä tutkimuksessa mainittiin myös, että Lavery et al. malli, sekä ”user action framework” eivät tarjonneet itsessään riittäviä ohjeita luokitteluun, joten tämän tutkimuksen järjestäjät rakensivat ohjeistuksen itse. Tästä syystä on hyvin mahdollista, että paremmilla ohjeilla, sekä sillä, että luokittelijat olisivat olleet kokeneita käyttämään näitä malleja oltaisiin saatu todella paljon parempia tuloksia. Näin ollen olisi hyvä tehdä vastaavanlaisia tutkimuksia ihmisillä, jotka ovat ehtineet saamaan riittävästi kokemusta käytettävyyden menetelmien luokittelusta ennen testin aloittamista, sekä tutkia lisää, miten luokittelu tulee käytännössä suoritaa.

González et al. [53] pyrkivät etsimään tehokasta tapaa hyödyntää käytettävyydesteistä saatua dataa. Työssä käytetään datan louhintatekniikoita, joilla pyritään löytämään kvalitatiivisesta käyttäjätutkimusdatasta toistuvia rakenteita tiettyssä kontekstissa. Työssä mainitaan myös, että löydettyjä tiettyyn kontekstiin liittyviä toistuvia käytettävyysongelmarakenteita voidaan hyödyntää myös toisissa samaan kontekstiin kuuluvissa tuotteissa. Tällöin voidaan helpommin myös tunnistaa käytettävyysongelmien syntymekanismit ja estää ne. Gonzales et al. [53] esitetään työssään kaksi tapaa joiden avulla löydetään käytettävyydatasta toistuvia rakenteita, eli assosiaatiosäännöt (association rules) ja päätöspuut (decision trees). Kun käytettävyydatasta löydetään selkeitä rakenteita voidaan niiden avulla muodostaa kokonaisuuksia eri käytettävyysongelmatyypeille ja näin saadaan luokiteltua käytettävyysongelmat. Kun ongelmat on luokiteltu nähdään, mitkä ongelmat nousevat tärkeimmiksi ongelmiksi ja niihin tulee kiinnittää erityistä huomiota.

Kun on kerätty raakaa käytettävyydataa sitä pyritään selkiyttämään lisäämällä mahdollisesti puuttuvia kohtia muodostaen selkeämpiä kokonaisuuksia, vähentämällä mahdollisia vääristäviä tekijöitä ja korjaamalla epäjohdonmukaisuuksia. Tämän jälkeen voidaan ottaa käyttöön González et al. esittämät datan louhintatekniikat.

Assosiaatiosäännöt (association rules) on tekniikka, jolla etsitään kiinnostavia yhteyksiä datassa olevien asioiden välillä. Tätä tekniikkaa on esimerkiksi käytetty analysoimaan minkälaisia yhteyksiä kaupassa asioivan asiakkaan ostosten välillä on. Näin kauppiat voivat luoda markkinointistrategioita, hyödyntäen tietoa siitä, mitkä tuotteet esiintyvät usein muiden tuotteiden kanssa asiakkaiden ostokoreissa. Ongelmaksi tässä menettelytavassa tulee kuitenkin hyvin helposti assosiaatiosääntöjen määrän kasvu liian suureksi suhteessa datan määrän kasvuun. Tämä toki voidaan ratkaista asettamalla kynnyksiarvoja sille miten vahva asioiden välinen yhteys tulee olla, jotta ne muodostavat toistuvan rakenteen. Toisaalta on myös kehitetty uusia todella tehokkaita algoritmeja kuten "appriori" ja "FPGROWTH" auttamaan assosiaatiosääntöjen muodostamisessa.

Eräs toinen vaihtoehto on päätöspuut (decision trees). Päätöspuut ovat vuokaavion kaltaisia puurakenteita, joissa jokainen solmukohta edustaa tutkimuksessa löytynyttä tärkeää asiaa ja lehtisolmut edustavat päämääriä. Kun ollaan saatu datasta kasa huomionarvioisia asioita, kuten käytettävyysongelmia, sekä päämääriä, eli miten käytettävyysongelmat vaikuttavat käyttäjään voidaan automaattisesti generoida päätöspuu. Useimmiten 2/3 löydetystä datasta käytetään päätöspuun rakentamiseen ja 1/3 käytetään päätöspuun toimivuuden testaamiseen. Kun käytettävyysongelmat on saatu sijoitettua päätöspuuhun ne muodostavat hierarkisen kokonaisisuuden. Näin ollen nähdään, mitkä käytettävyysongelmat ovat oleellisimpia ja mihin tulee puuttua. Esimerkkinä päätöspuusta on kuvassa (kuva 1.) näkyvä päätöspuu, joka yrittää määrittää ovatko lauantai-aamut sopivia tenniksen pelaamiselle.



Kuva 1. Päätöspuun esimerkki

Voidaan kulkea esimerkiksi polku Sääennuste → sateinen sää → tuulee → heikko tuuli → kyllä, eli pelataan tennistä tai polku Sääennuste → aurinkoinen sää → ilmankosteus → korkea → ei, eli ei pelata tennistä. Tämä on esimerkki yliyksinkertaistettu, mutta sen tarkoitus on vain selventää päätöspuun toimintaa.

Toisaalta Hornbæk et al [49] osoittavat, että mahdolliset ratkaisuehdotukset käytettävyyssongelmiin ovat hyödyllisempiä kuin lista tärkeimmistä käytettävyyssongelmista. Tämän tutkimuksen mukaan se johtuu siitä, että parannusehdotukset auttavat tuotekehittäjiä ymmärtämään käytettävyyssongelmia syvällisemmin. Syy tähän on se, että parannusehdotukset määrittelevät ongelmien luonteen selvemmin, sekä ne ovat konkreettisempia kuin pelkät käytettävyyssongelmat ja ne myös osoittavat, miksi on tärkeää puuttua näihin ongelmiin. Parannusehdotukset käytettävyyssongelmiin myös inspiroivat tuotekehittäjiä ajattelemaan vaihtoehtoisia ratkaisumalleja. Molichin ja kumppaneiden CUE-4 työ [11] myös painottaa käytettävyyssongelmien pohjalta tehtyjen parannusehdotusten merkitystä. Työ [49] toisaalta muistuttaa, että tuotekehittäjät haluavat silti saada myös listan käytettävyyssongelmista, sillä ne auttavat luokittelemaan heidän tekemiensä muutosten tärkeyttä, sekä toimivat todisteena muutosten hyödyllisyydestä. Lopuksi todetaan, että parannusehdotuksilla on suuri merkitys lopputuotteeseen. Olisi syytä tutkia tarkemmin, kuinka hyviä parannusehdotuksia tulisi rakentaa.

5 Yhteenveto

Jokaisella käytettävyytutkimusmenetelmällä on omat vahvat puolensa ja heikkoutensa. Myös käytettävyyssalan ammattilaiset ja tutkijat ovat sillä kannalla, että jokaisen käytettävyyssmenetelmän ominaisuudet määrittävät menetelmän tehokkuuden löytää tietyntyyllisiä käytettävyyso ongelmia. Näin ollen eri käytettävyydenarviointimenetelmiä tulisi yhdistellä ottaen huomioon eri menetelmien puutteet ja vahvuudet.

Kun verrataan käytettävyytutkimusta ja asiantuntija-arviointimenetelmiä huomataan niiden vastaavan eri kysymyksiin. Käytettävyyssmentauksella nähdään ainoastaan käytettävyyso ongelmien oireet ei syytä kun taas heuristisella arviolla saadaan nimenomaan käsitystä ongelmien syistä muttei nähdä kuinka ongelmat tarkalleen ilmenevät. Sen lisäksi heuristisella arviolla saadaan kattava, mutta pinnallinen käsitys käytettävyyso ongelmista, mutta käytettävyyssmentauksella saadaan hyvin tarkka käsitys joistakin tietyistä ongelma-alueista. Ääneen ajattelumenetelmä, sekä kognitiivinen läpikäynti taas pyrkivät vastaamaan kysymykseen miksi käyttäjä toimii tietyllä tavalla. Ääneen ajattelumenetelmä keskittyy suppeammin tiettyihin käytettävyyso ngelma-alueisiin tarjoten tarkkaa tietoa kun taas kognitiivnen läpikäynti tarjoaa laaja-alaisempaa mutta pintapuolisempaa tietoa käytettävyyso ongelmista.

Automaattisessa tiedonkeräyksessä taas pyritään ratkaisemaan useimmat käytettävyyssarvioon liittyvät ongelmat. Automaattisella tiedonkeräyksellä voidaan vähentää tarvetta käytettävyyssmentaukselle. Tällöin saadaan myös analysoitua huomattavasti suurempi määrä erilaisia käyttäjiä ja näin ollen löydetty käytettävyyso ngelmat ovat varmemmin sellaisia, jotka koskevat koko käyttäjäryhmää.

Jotta saavutetaan käytettävyytutkimuksessa paras mahdollinen tulos tulee hyödyntää monia eri menetelmiä yhdessä. On selvää, että tuotekehitysprosessissa aikaa vievämpiä menetelmiä, kuten käytettävyyssmentausta ei voida käyttää montaa kertaa ajanpuutteen ja kustannusten takia. Tästä syystä on välttämätöntä käyttää myös menetelmiä, joita voi käyttää iteratiivisesti tuotekehitysprosessin aikana.

Kun nopeilla iteratiivisilla menetelmillä, kuten heuristisella arviolla havaitaan riittävän vakavia käytettävyyso ongelmia voidaan ottaa käyttöön aikaa vievämpi ja tehokkaampi menetelmä kuten käytettävyyssmentaus ja keskittyä tarkastelemaan syvemmin heuristisella arviolla löydettyjä ongelma-alueita. Käytettävyyssmentauksen tukena on hyvä käyttää myös kyselyä varsinkin silloin jos käytettävyyssmentaus suoritetaan kenttätestauksena. Kenttätestaus tarjoaa yksityiskohtaista tietoa, siitä miten käyttäjä käyttää tuotetta oikeassa ympäristössä niin tällöin ne tuotteen osa-alueet, joihin keskitytään rajoittuvat pelkästään valittuihin tehtäviin. Tästä syystä kun sen tukena käytetään laajemman, mutta pinnallisemmän kuvan antavaa menetelmää kuten heuristista arviota, saadaan huomattavasti kattavampi kuva tuotteen kokonaiskäytettävyydestä. Koska käytettävyyssmentaus ja heuristinen arvio eivät juuri kerro mitään käyttäjän ”mental” malleista, niin samanaikaisesti on mahdollista suorittaa ääneen ajattelu -menetelmällä tapahtuva arvio, joka nimenomaan keskittyy siihen, miksi käyttäjät toimivat niin kuin toimivat.

Molichin ja kumppaneiden tutkimuksissa nousi esille useita mielenkiintoisia asioita. Yksi asia oli se, että kun useat eri käytettävyyssammattilaiset arvioivat samaa palvelua saadaan todella erilaisia käytettävyyso ongelmia, vaikka kyseessä olisi käytettävyyssmentaus, jossa käytetään samoja tehtäviä. Toinen asia oli se, että käytettävyydeltään hyvästäkin palvelusta löydetään huomattavasti enemmän ongelmia kuin niihin pystytään puuttumaan. Työn loppuisuus keskittyy

ottamaan kantaa näihin ongelmiin. Käyttäjien ja arvioijien määrän merkitys käytettävyydestä, järjestelmällisyyttä käytettävyysohjelmien löytämiseen, sekä tarvitaan tehokkaampia keinoja luokitella käytettävyysohjelmia.

Käytettävyyden tutkimus on vielä hyvin nuorta, joten käytössä olevat menetelmät ovat vielä jokseenkin keskeneräisiä. Silti käytettävyyden parantumisesta saadut hyödyt ovat sen verran merkittäviä, että tulevaisuudessa käytettävyyteen tullaan panostamaan entistä enemmän ja näin ollen myös käytettävyyden tutkimuksen merkitys kasvaa.

Lähteet

[1] Nielsen, J. Usability Engineering. San Diego, CA, USA: Academic Press, 1993. 362 s. ISBN 1-12-518406-9. Viitattu sivut 9, 17-20.

[2] Faulkner, Kristine. Usability Engineering. New York, USA: PALGRAVE, 2000. 244 s. ISBN 0-333-77321-7. Viitattu sivut 1 ja 2.

[3] SFS-EN ISO 9241-11, Näyttöpäätteillä tehtävän toimistotyön ergonomiset vaatimukset. Osa 11: Käytettävyyden määrittely ja arviointi (Ergonomic requirements for office work with visual display terminals (VDTs). Part 11: Guidance on usability (ISO 9241-11:1998)). Helsinki: Suomen standardisointiliitto. 1998. 44 s.

[4] Leventhal, L & Barnes, J. Usability engineering; process, products, and examples. New Jersey, USA: Pearson Education, Inc, 2007. 314 s. ISBN-13: 978-0-13-157008-5. ISBN-10: 0-13-157008-0. Viitattu sivu 16.

[5] Sutcliffe, A & Ryan, M & Doubleday, A & Springett, M. Model mismatch analysis: towards a deeper explanation of users' usability problems. City University, School of Informatics. Lontoo, Iso-Britania. 2000. 13 s.

[6] Anandhan, A & Dhanpani, S & Reza, H & Namasivayam, K. Web usability testing – CARE methodology. University of North Dakota, Department of Computer Science. Pohjois Dakota, USA 2006. 6 s.

[7] Atterer, R & Wnuk, M & Schmidt, A. Knowing the users' Every Move – User Activity Tracking for Website Usability Evaluation and Implicit Interaction. University of Munich, Media Informatics Group & Embedded Interaction Research Group. Munchen, Saksa. 2006. 10 s.

[8] Heimonen, Tomi. Käytettävyyden automaattinen arviointi. Tampereen teknillinen yliopisto, Tietojenkäsittelytieteen laitos. Tampere. 2005. 18 s.

[9] Ivory, M.Y & Hearts, M.A. The State of the Art in Automating Usability Evaluation of User Interfaces. University of California. Berkeley. Berkley, USA. 2001. 47 s.

[10] Molich, R & Ede, M.R & Kaasgaard, K & Karyukin, B. Comparative usability evaluation. Behaviour & Information Technology. 2004. 10 s.[viitattu 20.10.2008]. DOI: 10.1080/0144929032000173951.

[11] Molich, R & Dumas, J.S. Comparative usability evaluation (CUE-4). Behaviour & Information Technology. 2008. 20 s. [viitattu 20.10.2008]. DOI: 10.1080/01449290600959062

[12] Nielsen, J. 1994. Guerrilla HCI: Using Discount Usability Engineering to Penetrate the Intimidation Barrier. Jakob Nielsen's Alertbox.[viitattu 4.12.2008]. Saatavissa:http://www.useit.com/papers/guerrilla_hci.html.

- [13] Öörni, K. What do we know about usability evaluation? A critical view. University of Oulu. Department of Information Studies. 2003. 13 s. [viitattu 26.11.2008]. Saatavissa: <http://www.student.oulu.fi/~koorni/digilib2003.pdf>
- [14] Whittaker, S. & Terveen, L. & Nardi B. A. Let's stop pushing the envelope and start addressing it: A reference task agenda for HCI. Human-Computer Interaction. ATT Labs-Research. 2000. 33 s. [viitattu 26.11.2008]. DOI: 10.1207/S15327051HCI1523_2.
- [15] Hartson, H.R. & Andre, T.S. & Williges R.C. Criteria for Evaluating Usability Evaluation Methods. Virginia Tech. Department of Computer Science & Department of Industrial and Systems Engineering. Virginia, USA. 2001. 39 s. [viitattu 26.11.2008]. DOI 10.1207/S15327590IJHC1304_03.
- [16] Cordes, R.E. Task-Selection Bias: A Case for User-Defined Tasks. IBM Corporation. International Journal of Human-Computer Interaction 13(4), 411-419. Lawrence Erlbaum Associates, Inc. 2001. s 10.
- [17] Hertzum, M & Jacobsen N.E. Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. International Journal of Human-Computer Interaction 15(1), 183-204. Lawrence Erlbaum Associates, Inc. 2003. s 24.
- [18] Lewis, J.R. Evaluation of Procedures for Adjusting Problem Discovery Rates Estimated From Small Samples. IBM Corporation. [viitattu 26.11.2008]. DOI: 10.1207/S15327590IJHC1304_06
- [19] Jacobsen, N.E & Hertzum, M & John B.E. The Evaluator Effect in Usability Studies Problem Detection and Severity Judgments. Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting (Chicago, October 5-9, 1998), pp. 1336-1340. HFES, Santa Monica, CA. 1998. 5 s.
- [20] Kessner, M & Wood, J & Dillon R.F. & West R.L. On the reliability of Usability Testing. Carleton University, Department of Psychology. Ottawa, Canada. 2001. 2 s.
- [21] Matera, M & Rizzo, F & Carughi, G.T. Web Usability: Principles and Evaluation Methods. Politecnico di Milano, Dipartimento di Elettronica e Informazione. Milano, Italy. 2006. 39 s. [viitattu 26.11.2008]. Saatavissa: <http://www.webml.org/webml/upload/ent5/1/WebUsability-MateraEtAl.pdf>.
- [22] Benbunan-Fich, R. Using protocol analysis to evaluate usability of a commercial web site. Stillman School of Business, Seton Hall University. Computing and Decision Sciences Department. New Jersey, USA. 2001. 13 s.
- [23] Nielsen, J. 2005. Ten usability heuristics. ISSN 1548-5552. [viitattu 21.10.2008]. Saatavissa: http://www.useit.com/papers/heuristic/heuristic_list.html
- [24] Heuristisen arvioinnin muistilista -lyhyt versio. Kansalliskirjasto.[Viitattu 4.12.2008]. Saatavissa: <http://www.kansalliskirjasto.fi/attachments/513q0OxeV/51XFq9qmS/Files/CurrentFile/20050829>

_Heuristinen_arviointi.pdf

[25] Nielsen, J. How to Conduct a Heuristic Evaluation. [viitattu 21.10.2008]. Saatavissa: http://www.useit.com/papers/heuristic/heuristic_evaluation.html

[26] Zhang, Z & Basili, V & Shneiderman, B, Perspective-based Usability Inspection: An Empirical Validation of Efficacy. [Verkkolehti] Empirical Software Engineering. Vol 4, Number 1. 3.1999. S. 43-69. Springer Netherlands, 1999. DOI: 10.1023/A:1009803214692.

[27] Blandford, A.E & Hyde, J.K & Green T.R & Connel, I. Scoping Analytical Usability Evaluation Methods: A Case Study. University College London & Middlesex University & University of Leeds, Great-Britain.2008. 51 s. [viitattu 26.11.2008]. DOI: 10.1080/07370020802278254.

[28] Doubleday, A & Ryan, M & Springett, M & Sutcliffe, A. A Comparison of usability techniques for evaluating desig. City University, School of Informatics, Centre for HCI Design. London, UK. 1997. 10 s. [viitattu 26.11.2008]. Saatavissa: <http://portal.acm.org/citation.cfm?id=263583& amp;coll=>.

[29] Brooks, P. Adding value to usability testing. Teoksesta: Usability Inspection methods. S. 255-271. New York, USA. John Wiley & Sons, Inc, 1994. .ISBN:0-471-01877-5.

[30] Communications of the ACM. Volume 48, Number 1 (2005), Pages 71-74. Usability Engineering for Software Developers. 2005. 4 s. [viitattu 26.11.2008]. Saatavissa: <http://portal.acm.org/citation.cfm?doid=1039539.1039541>

[31] Scholtz, J. Usability Evaluation. National Institute of Standards and Technology. 2004. 8 s. [viitattu 26.11.2008]. Saatavissa: http://www.itl.nist.gov/iad/IADpapers/2004/Usability%20Evaluation_rev1.pdf

[32] Andre, T.S. & Hartson, H.R. & Williges, R.C.Determining the Effectiveness of the Usability Problem Inspector: A Theory-Based Model and Tool for Finding Usability Problems. Virginia Polytechnic Institute and State University, Balcksburg, Virginia, USA. 2003. 286 s. [viitattu 26.11.2008]. ISSN: 00187208. Saatavissa: <http://scholar.lib.vt.edu/theses/available/etd-04122000-09440030/unrestricted/andre.pdf>

[33] Ramey, J & Boren, T & Cuddihy, E & Dumas, J & Guan, Z & van den Haak, M.J & De Jong, M.D.T. Does Think Aloud Work? How Do We Know? Teoksessa: CHI 2006 extended abstracts on Human factors in computing systems. Montréal, Québec, Canada. 22–27.4.2006. New York, USA. ACM 1-59593-298-4/06/0004, 2006. S. 45-48. [viitattu 4.12.2008]. Saatavissa: <http://portal.acm.org/citation.cfm?id=1125451.1125464>. ISBN:1-59593-298-4.

[34] Fu, L. & Salvendy, G. & Turley, L. Effectiveness of user testing and heuristic evaluation as a function of performance classification. BEHAVIOUR & INFORMATION TECHNOLOGY. [Verkkolehti]. VOL. 21, NO. 2. 2003. S. 137-143. [Viitattu 4.12.2008]. DOI: 10.1080/0269905011011368 8.

[35] Rohrer, C. When to Use Which User Experience Research Methods. Jakob Nielsen's Alertbox. 10.6.2008. [viitattu 4.12.2008]. Saatavissa: <http://www.useit.com/alertbox/user->

research-methods.html.ISSN 1548-5552.

[36] Marty, P.F. & Twidale M.B. Extreme Discount Usability Engineering. Florida State University, College of Information, Tallahassee, Florida, USA & University of Illinois, Graduate School of Library and Information Science, Champaign, Illinois, USA. 2005. 22 s. [Viitattu 4.12.2008]. Saatavissa: <http://www.isrl.uiuc.edu/~twidale/pubs/ExtremeDiscUETechReport.pdf>.

[37] Kantner, L. & Rosenbaum, S. Usability Studies of WWW Sites Heuristic Evaluation vs Laboratory Testing. Proceedings of the 15th annual international conference on Computer documentation. New York, USA. ACM, 1997. S. 153-160. ISBN:0-89791-861-4.

[38] NIELSEN, J. Why you only need to test with 5 users. Jakob Nielsen's Alertbox. 19.11.2000. [Viitattu 4.12.2008]. Saatavissa: <http://www.useit.com/alertbox/20000319.html> <http://www.useit.com/alertbox/20000319.html>. ISSN 1548-5552.

[39] Faulkner, L. Beyond the five-user assumption: Benefits of increased sample sizes in usability testing. Behavior Research Methods, Instruments, & Computers, Volume 35, Number 3, 1.8.2003. S. 379-383. [Viitattu 4.12.2008]. Saatavissa: <http://docserver.ingentaconnect.com/deliver/connect/psocpubs/07433808/v35n3/s5.pdf?expires=1228417063&id=47575753&titleid=1013&acname=Helsinki+University+Of+Technology&checksum=C312006B2369B0FF9D3451ABF21CFE4D>.

[40] Cockton, G & Woolrych, A. Sale Must End: Should Discount Methods be Cleared off HCI's Shelves? Vol 9, Issue 5, 12.2002. S. 13-18.[viitattu 4.12.2008]. Saatavissa: <http://osiris.sunderland.ac.uk/~cs0awo/interact%202002.pdf>. ISSN:1072-5520.

[41] Lindgaard, G. & Chattratichart, J. Usability Testing What Have We Overlooked. Conference on Human Factors in Computing Systems. Proceedings of the SIGCHI conference on Human factors in computing systems. New York, USA. ACM, 2007. S. 1415-1424. ISBN:978-1-59593-593-9.

[42] Alsharnari, M & Mayhew, P. Task Design: Its Impact on Usability Testing. The Third International Conference on Internet and Web Applications and Services. DOI 10.1109/ICIW.2008.20.

[43] Lewis, J.R. Introduction Current Issues in Usability Evaluation. International Journal of Human-Computer Interaction. [Verkkolehti]. Vol 13, Issue 4. 11.2001. S. 343-349. [Viitattu 4.12.2008]. DOI: 10.1207/S15327590IJHC1304_01.

[44] Chattratichart, J. & Lindgaard, G. A Comparative Evaluation of Heuristic-Based Usability Inspection Methods. Conference on Human Factors in Computing Systems. Chi '08 extended abstracts on Human factors in computing systems. New York, USA. ACM, 2008. S. 2213-2220. ISBN:978-1-60558-012-X.

[45] Kort, J. & de Poot, H. Usage Analysis Combining Logging and Qualitative Methods. Conference on Human Factors in Computing Systems. CHI '05 extended abstracts on Human factors in computing systems. Portland, Oregon, USA. S. 2121-2122. New York, USA. ACM, 2005. ISBN:1-59593-002-7.

- [46] Berners-Lee, T. What is the future of the internet? [Haastattelu]. BBC, 9.7.2008. [Viitattu 4.12.2008]. Kesto 8 min 36 s. Saatavissa: http://news.bbc.co.uk/today/hi/today/newsid_7496000/7496976.stm
- [47] Dicks, R.S. Mis-Usability: On the Uses and Misuses of Usability Testing. Proceedings of the 20th annual international conference on Computer documentation. Toronto, Ontario, Canada. S. 26-30. New York, USA. ACM, 2002. ISBN:1-58113-543-2.
- [48] Hornbæk, K & Frøkjær, E. Comparison of techniques for matching of usability problem descriptions. University of Copenhagen. Department of Computer Science. Copenhagen, Denmark. 2008. 10 s. DOI:10.1016/j.intcom.2008.08.005.
- [49] Hornbæk, K & Frøkjær, E. Comparing Usability Problems and Redesign Proposals as Input to Practical Systems Development. Conference on Human Factors in Computing Systems. Proceedings of the SIGCHI conference on Human factors in computing systems. Portland, Oregon, USA. S. 391-400. New York, USA. ACM, 2005. ISBN:1-58113-998-5.
- [51] Redish, J & Bias, R.G. & Bailey, R. & Molich, R. & Dumas, J. Spool, J.M. Usability in Practice: Formative Usability Evaluations - Evolution and Revolution. Conference on Human Factors in Computing Systems. CHI '02 extended abstracts on Human factors in computing systems. Minneapolis, Minnesota, USA. S. 885-890. New York, USA. ACM, 2002. ISBN:1-58113-454-1.
- [52] Labroche, N. & Lesot, M-J. & Yaffi, L. A New Web Usage Mining and Visualization Tool. Tools with Artificial Intelligence, 2007. ICTAI 2007. 19th IEEE International Conference on Tools with Artificial Intelligence. S. 321-328. DOI:10.1109/ICTAI.2007.102.
- [53] González, M.P. & Lorés, J. & Granollers, A. Enhancing usability testing through datamining techniques: A novel approach to detecting usability problem patterns for a context of use. National Council of Scientific and Technical Research (CONICET), Buenos Aires, Argentina & Universidad Nacional del Sur, Department of Computer Science and Engineering, Visualization and Graphics Research Laboratory, VyGLab, Bahía Blanca, Argentina. 2008. 22 s. DOI:10.1016/j.infsof.2007.06.001.