

TPÜ HAAPSALU KOLLEDŽ
Infotehnoloogia osakond

Rene Saarsoo

**STANDARDID
JA
EESTI VEEBIMAASTIKU OLUKORD
NENDE OSAS**

Proseminaritöö uurimisprojekt

Juhendaja: mag. Jaagup Kippar

Haapsalu 2005

SISUKORD

Sissejuhatus.....	3
1.Probleemi kirjeldus.....	5
1.1.Uuringud Eestis.....	5
1.2.Uuringud mujal.....	5
1.3.Ülesanne.....	6
2.Materjal ja metoodika.....	7
2.1.Probleemid varasemate uurimustega.....	7
2.2.Meetmed uuringu parandamiseks.....	7
2.3.Töös kasutatavad vahendid.....	8
Kokkuvõte.....	9
Viited.....	10

SISSEJUHATUS

Veeb on noor meedia. Meedia, millel puuduvad selged piirid ja kammitsad. Selline vabadus loob meile ühelt poolt suured võimalused oma mõtete ja ideede teostamiseks, ent tõeline vabadus ei saa olla anarhia, kus igaüks lihtsalt teeb seda, mis tal pähe tuleb ja ei arvesta mitte põrmugi teiste huvidega, hoomamata sealjuures, et need teiste huvid võivad tegelikult olla ka tema enese huvid.

Kui Veebi looja Tim Berners Lee 1994. aastal asutas Ülemaailmse Veebikonsortsiumi (*World Wide Web Consortium – W3C*) (Jacobs 2000), siis oli ideeks tuua korda ja tasakaalu sellesse ettearvamatusse hüpertextimaailma – nii et hundid oleks söönud ja ka lambad terved – seda ideed kannab W3C tegevus kuni tänaseni.

W3C tegevuse põhisuunaks oli ja on mitmesuguste standardite (tuntud kui W3C soovituselised – *recommendations*) väljatöötamine. Kuna veebi põhiliseks formaadiks on HTML (*Hypertext Markup Language – hüpertexti märgenduskeel*), siis on kõige suuremad kired keerelnud just selle märgenduskeele ümber. Alates esimesest veebilehitsejast ehk brauserist nimega World Wide Web hakkasid paljud brauseritootjad välja mõtlema mitmesuguseid täiendusi HTML-keelele, mis algselt ei võimaldanud mitte midagi muud kui lihtsalt teksti struktureerida. Mosaic tuli välja võimalusega kasutada lisaks tekstile ka pilte, Netscape tegi võimalikuks veebilehtede skriptimise luues JavaScript´i nimelise programmeerimiskeele jne. Kõik see viis välja selleni, mida tuntakse kui brauserite sõda (*browser wars*), mis toimus Microsoft Internet Explorer´i (IE) ja Netscape Navigator´i vahel ning millest väljus võitjana IE.

Ka brauserisõdade aegu olid juba olemas W3C loodud standardid, aga brauseritootjad ei hoolinud neist kas üldse mitte või proovisid nende alusel luua teatud sorti oma nägemuse. Tänapäevaks on siiski selgeks saanud, et ilma ühtse standardita me kaugele ei jõua. Ühegi veebiarendaja jaoks pole mõistlik kasutada mingi kindla brauseri spetsiifilisi lahendusi,

sest tehes nõnda on oht kaotada teisi brausereid kasutavad kliendid. Brauseritootjad ei võistle enam mitte selles, milliseid uue funktsiooniga HTML-i elemente nad välja mõtlevad vaid selles, kuid võrd hästi keegi neist toetab W3C standardeid (isegi selles vallas paljukirjutud Microsoft on teinud pingutusi selles vallas ja paistab jätkavat (Microsoft... 2004)). Standardsete, kõigis moodsates brauserites ühtviisi töötavate (ja vanemates kenasti degradeeruvate) lahenduste loomine, on hetkel lihtsam kui kunagi varem.

Ent kas ka veebidisainerid ja programmeerijad seda endale teadvustavad? Üldine praktika näitab, et enamasti mitte. Endiselt on levinuimaks brauseriks IE, mis saab automaatselt iga uue Windowsi kasutaja kaaslaselks ning milleks hoolida ülejäänud marginaalsest hulgast alternatiivsete brauserite kasutajatest – tühi töö ja vaev?

Viimaste aastate jooksul on aga alternatiivsete brauserite kasutamine suurenenud üha ning IE viimane versioon 6.0, millel vanust juba mitmed aastad, mõjub nende noorte tegijate kõrval (kelle viimane väljalase pole ühelgi juhul vanem kui aasta), tõelise vanurina. Ning mitte ainult brauserid pole põhjus kasutamaks standardset ja korrektset HTML-i, ka otsingurobotid, Google'iga eesotsas, suudavad lehekülge paremini indekseerida. Kuigi semantiline HTML otseselt otsingumootorite hinnangut lehest ei mõjuta (Meyer 2004).

Suure ja olulise tõuke standardite kasutamise suunas on andnud ka mitmete riikide valitsused, eesotsas USA-ga. Nimelt on seadustes, mis määravad tingimused riigiasutuste kodulehekülgedele küllaltki kõrged nõudmised ligipääsetavuse osas, kusjuures soovituslikuks meetodiks ligipääsetavuse tagamiseks on korrektne HTML-i ja CSS-i (*Cascading Style Sheets* – kaskaadivad stiililehed) kasutus. Hoolimata seadustest pole olukord nendes riikides veel kaugeltki kiiduväärne, ehkki see paraneb järjest. Eestis puuduvad aga vastavasisulised seadused üleüldse.

1. PROBLEEMI KIRJELDUS

1.1. Uuringud Eestis

Vaadates vaid mõningaid populaarsemaid veebilehekülgi ei saa me teha oluliselt täpsemaid järeldusi kui: „Enamik Eesti veebilehti veebistandardeid ei järgi.” Konkreetsemad uuringud antud valdkonnas on aga puudulikud. On küll Valdo Prausti poolt 2002. aasta lõpus teostatud „Eesti riigiasutuste veebilehtede käideldavustest” (Praust 2002, 2003), mis aga käsitleb vaid ligipääsetavausega seotud probleeme ning sedagi vaid mõnekümnel riigiasutuse leheküljel.

Uuringuid, mis aga käsitleksid veebilehtede standarditele vastavust (valideerivust) pole aga teostatud üldse, va muidugi pisemat sorti testimised, mida üks või teine eraisik on oma puht isiklikust huvist teinud (näiteks minu poolt korraldatud uuring Eesti veebidisainifirmade avalehtede HTML-i valideerimise kohta).

1.2. Uuringud mujal

Mujal maailmas on küll mõningaid uuringuid korraldatud, kuid enamasti on need olnud kas liialt väikese mahuga või liialt pealispinnalised. Järgnevalt toon ära uurimused, mille kohta on õnnestunud informatsiooni hankida; täpsemad arvanded on ära toodud tabelis 1.

- a) 1997 Thomas Dowling „Validating HTML” (tegemist on artikli nimega, tegelik uurimus on läbi viidud enne artikli kirjutamist). Teostati 624 Põhja-Ameerika raamatukogu kodulehekülge valideerimine HTML 3.2 dokumenditüübi alusel.
- b) 2003 Greg Lanier „Universal web design: a survey of web accessibility and usability”. Uuriti 41 väljavalitud lehekülge erinevatest valdkondadest (otsingumootorid, valitsusasutused, veebistandarditega seotud leheküljed, ...). Uuriti

valideerivust ja vastavust Section 508 ning WCAG 1.0 ligipääsetavusnõuetele ning hinnati HTML-leheküljel oleva teksti ja selle esitamiseks vajamineva koodi osakaalu.

- c) 2. märts 2004 Soren Johannessen „The use of W3C standards in Denmark”. Uuriti 2033 Taani valitsus-, rahvus- ja munitsipaalasutuse lehekülje valideerumist. Lisaks tuvastati kui paljudel lehekülgedel on olemas dokumenditüübi deklaratsioon ning erinevate dokumenditüüpide osakaalud.
- d) 16. märts 2004 Soren Johannessen „Danish companies and the use of W3C standards”. Täpselt samas vormis uuring, mis eelmine. Sel korral teostatud 250 Taani äriettevõtte koduleheküljel.

<i>Uurimus</i>	<i>Lehekülgede arv</i>	<i>Valideerivaid lehti</i>	<i>Protsent</i>	<i>Keskmine vigade arv</i>
A	624	24	3.80%	20
B	41	7	17.00%	109
C	2033	62	3.10%	
D	250	1	0.40%	

Tabel 1. Erinevate uurimuste tulemuste võrdlus.

1.3.Ülesanne

Kuna puudub ülevaade sellest, milline on olukord veebistandardite järgimise ja seega ka lehekülgede kvaliteedi osas, siis oleks tarvis teostada uuring mis käsitleks Eesti veebilehekülgede valideerivust vastavalt W3C validaatorile, mis olekski antud proseminaritöö teemaks.

2. MATERJAL JA METOODIKA

2.1. Probleemid varasemate uurimustega

Eelmises peatükis loetletud HTML-i valideerivuse uuringutel ilmnes mitmeid probleeme:

- a) Paljusid lehekülgi ei saanud valideerida, kuna puudus dokumenditüübi deklaratsioon.
- b) Paljusid lehekülgi ei saanud valideerida, kuna neil kas puudus kasutatava kooditabeli määratlus või oli see vale.
- c) Valideerimisel mõõdeti vaid vigade arvu (kui sedagi), mis aga iseenesest pole eriti heaks kriteeriumiks hindamaks lehe kvaliteeti. Põhjuseid on kolm:
 1. Kõigest üks tilluke viga dokumendi alguses võib tekitada tuhandettesse (reeglina ühte-kahte tüüpi) vigadesse ulatuva kaskaadi.
 2. Enamik mittejuhuslikke vigu HTML-is reeglina korduvad (näiteks kui ei teata, et piltidele tuleb kindlasti määrata alternatiivtekst, siis arvatavasti ei tehta seda ühegi pildi puhul, aga pilte võib leheküljel olla sadu).
 3. Osad validaatori antavad veateated pole mitte vead vaid viited eelmise vea potentsiaalsele tekkekohale.

2.2. Meetmed uuringu parandamiseks

Vältimaks nimetatud probleeme tuleks kavandatavas uurimuses võtta kasutusele järgmised meetmed:

- a) Dokumenditüübi puudumise probleemi on lahendanud hiljutine W3C validaatori uuendus, mis kasutab dokumenditüübi puudumisel *HTML 4.01 Transitional* dokumenditüüpi, mis on kõige laiahaardelisem ja ühtlasi kõige leebem dokumenditüüp. Samas peab hoolitsema selle eest, et hilisemas infotöötluses oleks

võimalik eristada millistel lehtedel dokumenditüübi määrang puudus ja millistel oli see olemas.

- b) Kooditabeli määrangu puudumisel tuleks pakkuda mõnda ISO 8859 perekonda kuuluvat tabelit (validaator neil vahet ei tee, sest programmi jaoks pole oluline, kas on kirjutatud „žee“ või „beee“ ja enamasti on veebilehed just mõnes sellises ühebaadises vormingus, mis aga läheb vastuollu W3C validaatori poolt vaikimisi kasutatava mitmebaadise UTF-8 kodeeringuga). Mõistagi tuleb hilisemast kodeeringute kasutamise analüüsist välja jätta need leheküljed, millel kodeering algselt määratletud polnud.
- c) Valideerimisel tuleks koguda infot erinevate vigade esinemissageduste kohta, mis annaks võimaluse analüüsida lisaks vigade hulgale lehel ka *erinevate* vigade hulka ning enimlevinud vigu. Samuti saab sel teel validaatori veateadete hulgast „mitte päris vead“ välja praakida.

Töö maht peaks olema võimalikult suur – plaan on käia validaatoriga üle kõik(!) www-ga algavad ja .ee-ga lõppevad leheküljed. Vaatluse alla tulevad siiski vaid avalehed järgneval kolmel põhjusel:

- a) Avalehe kallal lehekülje loojad reeglina kõige rohkem vaeva näinud.
- b) Üheainsa parajalt mahuka kodulehekülje kõigi alamlehtede läbikäimine võib osutada juba omaette hiigeltööks.
- c) Ka kõik teised uuringud on keskendunud vaid avalehtedele.

2.3. Töös kasutatavad vahendid

- a) W3C Validaator (*W3C Markup Validation Service*) - valideerimise teostamiseks. (Aadressil: <http://validator.w3.org>).
- b) Omaloodud programm – suure hulga lehekülgede validaatorile etteandmise ja andmete kogumise automatiseerimiseks.
- c) Tabelitöötlusprogramm hilisemaks andmete statistiliseks analüüsimiseks.

KOKKUVÕTE

Uurimustöö „Standardid ja Eesti veebimaastiku olukord nende osas” peaks muu hulgas andma vastused järgmistele küsimustele:

- a) Kui suur osa Eesti veebilehekülgedest valideerib?
- b) Millised on levinuimad vead veebilehtede koodis?
- c) Milline on erinevate dokumenditüüpide kasutamise osakaal?
- d) Milline on erinevate kooditabelite kasutamise osakaal?

Vastava mass-valideerimise programmi loomine peaks hülbustama sarnaste uuringute kordamise tulevikus ja teistel lehekülgedel.

VIITED

- Dowling, T. (1997). *Validating HTML*. <http://silver.ohiolink.edu/tdowling/validation.html>
- Jacobs, I. (2000). *About the World Wide Web Consortium (W3C): Background*.
<http://www.w3.org/Consortium/#background>
- Johannessen, S. (2004). *Danish companies and the use of W3C standards*.
<http://www.ae35-unit.dk/standard/danishcompanies.html>
- Johannessen, S. (2004). *The use of W3C standards in Denmark*.
<http://www.ae35-unit.dk/standard/english.html>
- Lanier, G. (2003). *Universal web design: a survey of web accessibility and usability*.
<http://wwwx.cs.unc.edu/Research/assist/et/projects/universal/>
- Meyer, E. (2004). *SES Chicago Report*. [2004, september 18]
<http://www.meyerweb.com/eric/thoughts/2004/12/18/ses-chicago-report/>
- Praust, V. (2002). *Eesti riigiasutuste veebilehtede käideldavustest*.
<http://www.ria.ee/atp/print.html?id=480>
- Praust, V. (2003). Eesti riigiasutuste veebilehed *Arvutimaailm* 2, 38-41.
- Preview of Web Development with Visual Studio 2005: Better Standards Support*. (2004)
Microsoft ASP.NET 2.0 and Visual Studio .NET „Whidbey“ Team.
<http://www.asp.net/whidbey/whitepapers/VSWhidbeyOverview.aspx?tabindex=0&tabid=1>
- Saarsoo, R. (2004). *Veebitegijad, mis teil on enda kaitseks öelda?* Postitus Planet'i foorumis. [2004, märts 4]. <http://forum.planet.ee/showthread.php?t=47192>