

TLÜ HAAPSALU KOLLEDŽ
Infotehnoloogia osakond

Rene Saarsoo

VEEBILEHTEDE KODEERIMISPRAKTIKAD

Bakalaureusetöö uurimisprojekt

Juhendaja: mag. Jaagup Kippar

Haapsalu 2005

1. SISUKORD

Sissejuhatus.....	3
1 Probleemi kirjeldus.....	5
1.1 Kas valideeruvate lehtede osakaal on suurenenud?.....	5
1.2 Kuidas on muutunud heade ja halbade praktikate suhe?.....	5
1.3 Veebilehtede kodeerimispraktikaid on vähe uuritud.....	6
1 Küsimused, millele soovime vastust.....	7
2.1 (X)HTML.....	7
2.2 CSS.....	7
2.3 JavaScript.....	8
2.4 Flash.....	8
2.5 Java.....	8
2.6 Uudisvood.....	9
2 metoodika.....	10
2.1 Valim.....	10
2.2 Programmi pool.....	10
Kokkuvõte.....	11
Viited.....	12

SISSEJUHATUS

Olles läbi viinud analüüsi 83-l Austraalia veebilehel, teatas John Allsopp Sydney's toimunud konverentsil „Web Essentials 05“:

- *Ma ei taha enam kunagi näha Firefox'i tööriistariba.*
- *Ma ei taha enam kunagi näha W3C validaatorite või „Cynthia Says“ tulemusi.*
- *Ma ei taha vaadata mitte ühegi tabelitel baseeruva lehekülje lähtekoodi.*

Kuid on lihtsalt häbiväärne, et nende lehekülgede autorid ei paista seda samuti teha tahtvat.

Igaüks, kes on vähegi kursis World Wide Web Consortium'i (W3C) HTML¹-i ja CSS²-i standarditega³, teab ka fakti, et veebilehti, mis nendest kinni peavad, on vaid käputäis. Seega, milleks meile üldse need standardid, kui enamik lehti töötab suurepäraselt ilma nendetagi? Milleks kulutada oma raha ja aega täitmaks mitmesuguseid eeskirju, kui me võiksime oma ressursse selle asemel hoopis kuskil paremas kohas rakendada?

Mida te arvate, kui suured peavad olema ühe veebidisaineri teadmised, et saada aru, mis on valesti järgmistes koodilõikudes:

```
</head><center><body bgcolor="#ffffff"...
```

```
</CENTER></BODY></HTML></form>
```

```
</font></font></font></font></font></font>
```

1 *Hypertext Markup Language* – hüperteksti ülestähendamise keel
2 *Cascading Style Sheets* – kaskaaduvad stiililehed
3 Ametlikult on tegemist siiski soovitusetega (*recommendations*).

Loogiliselt võttes peaks tegu olema ühe algaja veebimeistri näpukaga – ent oleks see vaid nii. Ülalolevad koodijupid pärinevad Boeing'i ametlikult koduleheküljelt. Aga Boeing toodab ju lennukeid! Kas Boeing'i lennukid on niisama vigased kui nende veebileht? Sugugi mitte – lennukid on esmaklassilised, läbinud kõikvõimalikud peened testid ja turvakontrollid. Boeing'i käive oli 2004. aastal 54 miljardit dollarit! Nad võiksid endale soovi korral palgata kõige esmaklassilisema veebimeistri maailmas. Neil pole vähimatki vabandust serverimaks vigase koodiga veebilehti. Aga just seda nad teevad.

Paljude veebimeistrite arvates ei tekita vigase HTML-i kirjutamine mingeid probleeme. Aga läbi tarkvaraarenduse ajaloo on teada palju taolisi halbu praktikaid, mida pole just kergesti maha tahtud jätta: GOTO käsud, globaalsed muutujad, kommentaaride puudumine koodist... Ka veebilehtede lohaka kirjutamise osas peab tekkima üldtunnustatud arusaam, et tegu on halva praktikaga.

Veebistandardite aktivistid on teinud juba aastaid „kihutustööd“, ning mitte edutult. Tänapäevaks kuuluvad standardipõhiste lehtede hulka juba ka sellised „tegijad“ nagu msn.com ja slashdot.org.

Ent kõik see edu võib olla näiline, sest samal ajal, kui on kasvanud veebistandarditega kooskõlas olevate lehtede hulk, on kasvanud ka vigaste lehtede arv. Ning seega võib kokkuvõttes olla nõnda, et standardeid järgivate lehekülgede osakaal on hoopistükkis langenud.

Aga see ei saa ju nii olla! Veebistandardite alaseid uudisgrupe, liste ja foorumeid on tekkinud nagu seeni pärast vihma. Palju mainekad saidid (nagu näiteks Macromedia, AOL⁴ ja MSN) on läinud üle CSS-i põhisele paigutusele. Olukord *peab* olema paranenud! Kuid keegi pole viinud läbi uuringut, mis ütleks, kui palju olukord ikkagi paranenud on?

Sellise uuringu läbiviimine ongi selle bakalaureusetöö uurimisprojekti esmaseks teemaks.

4 *America Online*

1 PROBLEEMI KIRJELDUS

1.1 Kas valideeruvate⁵ lehtede osakaal on suurenenud?

Minu enda hiljaaegu läbi viidud uuringud Eesti veebilehtedel andsid tulemuseks, et ehkki poole aastaga oli valideeruvate lehtede arv kasvanud (4 lehe võrra), oli mäejagu rohkem kasvanud mittevalideeruvate lehekülgede hulk (tervel 628 lehe jagu!), mistõttu valideeruvate lehtede osakaal oli hoopistükkis langenud 2,22%-lt 2,17%-le.

Kuna aga minu uuringutes kasutatud valim oli kaunis väike (ligikaudu 20 000), siis pole tulemused piisavalt usaldusväärsed, et hüpoteesi kinnitada. Pealegi, ei saa vaid Eesti lehtedelt kogutud andmete alusel teha üldistusi kogu maailma kohta.

Aastal 2001 viis norrakas Dagfinn Parnas oma magistrیتöö käigus läbi suuremahulise uuringu. Tema valimis oli 2,4 miljonit URL⁶-i, mis pärinesid vabatahtlike poolt koostatud veebikataloogist „Open Directory Project“. Kontrollides kõiki neid lehti HTML-i validaatoriga sai ta tulemuseks, et vaid 0,71% kõigist lehtedest valideerusid.

Kui Parnase poolt läbi viidud uuringut nüüd, 5 aastat hiljem, korrata – peaksimegi saama anda üpris täpse hinnangu sellele, kui palju (ja kas üldse) on valideeruvate lehtede osakaal suurenenud.

1.2 Kuidas on muutunud heade ja halbade praktikate suhe?

Kui valideerumise hindamise läbiviimine programmi poolt on võrdlemisi lihtne – leht kas

5 Siin ja edaspidi mõistetakse valideerumise all vastavust formaalsetele standarditele, nagu näiteks veebilehe HTML-koodi vastavus HTML-i dokumenditüübi deklaratsioonile. Valideerimise all mõeldakse protsessi, kus seda vastavust kontrollitakse – reeglina mõnda spetsiaalset programmi ehk validaatorit kasutades (näiteks W3C HTML-i ja CSS-i validaatorid).

6 *Uniform Resource Locator* – tavamõistes internetiaadress.

valideerub või mitte – siis näiteks selle hindamine, kas lehekülje kujundamiseks on kasutatud tabeleid, on juba komplitseeritud. Vaadates asja teise külje pealt – nende lehtede osakaal, kes näitavad oma esiküljel tabulaarset infot, peaks olema konstantne suurus, ning seetõttu ei tohiks häirida, kui meid huvitab see, kas tabelite väärkasutus on kasvanud või kahanenud.

Hoopis suuremaks probleemiks osutub siinkohal see, et puuduvad varasemad andmed, millega kogutud tulemusi võrrelda. Siinkohal jõuamegi probleemini, mis ajendas kirjutama uurimust pealkirja „Veebilehtede kodeerimispraktikad“ all.

1.3 Veebilehtede kodeerimispraktikaid on vähe uuritud

Enamik uuringuid selles vallas on olnud:

1. väga väikesemahulised (alla tuhande lehe),
2. keskendunud vaid vaid mõnele kitsale lihtsalt automatiseeritavale valdkonnale (valideerumine) ja/või
3. vaadelnud lehekülgi vaid mõnest kitsast regioonist (Eesti).

Seetõttu ongi minu töö eesmärgiks viia läbi põhjalikum uuring, mis ehkki ei anna hetkel võimalust öelda, kuid võrd on olukord veebis aja jooksul muutunud, loob eeldused tegemaks neid võrdlusi tulevikus. Mis aga veelgi tähtsam – loob ülevaate sellest, milline on olukord nüüd ja praegu.

1 KÜSIMUSED, MILLELE SOOVIME VASTUST

Järgnevalt on üles loetletud küsimusi, millele antud uuring peaks vastuse andma. Kindlasti on lisaks nendele veel mitmeid küsimusi millele see uuring vastuse annab, küllap on nende seas aga ka neid probleeme, mis vastuseta jäävad.

2.1 (X)HTML

- Kui paljud leheküljed valideeruvad?
- Millised on levinuimad dokumentitüübid?
- Millised on levinuimad kodeeringud?
- Millised on levinuimad failitüübid (*content-type*)?
- Millised on levinuimad elemendid?
- Millised on levinuimad atribuudid?
- Kui palju on tühja ALT-tekstiga pilte?
- Millised on levinuimad pildiformaadid?
- Milline on lehe koodi ja sisu mahtude suhe?
- Milline on levinuim alamlehtede arv (linkide arv) pealehel?
- Kui paljud lehed lingivad W3C validaatorit vms?
- Kui paljudel lehtedel on XML-i deklaratsioon?
- ...

2.2 CSS

- Kui paljud leheküljed kasutavad CSS-i?
- Kuivõrd kasutatakse *inline*, *embedded* ja *external* CSS-i?
- Kui suur on väliste CSS-i failide arv? Alternatiivsete stiililehtede kasutamine?

- Milliseid meediatüüpe kasutatakse?
- Milliseid selektoreid kasutatakse?
- Millised on levinuimad CSS-i atribuudid?
- Millised on levinuimad atribuutide väärtused?
- Milliseid on levinuimad mõõtühikud?
- Millised on levinuimad pildiformaadid CSS-is?
- Kui paljude lehtede CSS valideerub?
- ...

2.3 JavaScript

- Kuivõrd kasutatakse *inline*, *embedded* ja *external* JavaScripti?
- Kui suur on väliste JavaScripti failide arv?
- Kui palju kasutatakse DOM-i?
- ...

2.4 Flash

- Kui paljud lehed kasutavad Flashi?
- Milliseid HTML-i elemente kasutatakse Flashi sidumiseks?
- ...

2.5 Java

- Kui paljud lehed kasutavad Javat?
- Milliseid HTML-i elemente kasutatakse Java sidumiseks?
- ...

2.6 Uudisvood

- Kui paljud leheküljed pakuvad RSS ja/või Atom voogusid?
- ...

2 METOODIKA

2.1 Valim

Valimiks on „Open Directory Project“ leheküljelt alla laetav RDF, kuna seda kasutas ka oma uuringus ka Dagfinn Parnas. Mõnel teisel moel koostatud valimi põhjal saadud andmed ei pruugi aga olla Parnas'e tulemustega võrreldavad. Täpse valimi suuruse ja kuju peab selgitama analüüs.

2.2 Programmi pool

Veebilehtede koodi analüüsimiseks on plaanis kasutada võimalikult palju olemasolevaid teeke. Näiteks Perli jaoks on olemas HTML-Tree, CSS SAC, LibWWW jpm. Vabalt saab veebist alla laadida ka W3C HTML-i ja CSS-i validaatorid.

Kogutud andmed on kavas salvestada relatsioonilisse andmebaasi edasiseks mugavamaks töötamiseks. Teine märksa mõjuvam põhjus selleks on see, et andmeid saab hinnanguliselt olema liialt palju selleks, et nende töötlemine väljaspool andmebaasi oleks mõeldamatu.

KOKKUVÕTE

Käesoleva bakalaureusetöö eesmärgiks on uurida levinuimaid praktikaid veebilehtede kodeerimisel. See hõlmab endas selliseid tehnoloogiaid nagu HTML, CSS, JavaScript, Flash ja Java. Peamiselt keskendutakse sellele, et selgitada välja, kuivõrd levinud ühed või teised praktikad veebis on, ning võimalust mööda ka seda, kuidas on nende osakaal aja jooksul muutunud.

Töö peaks anda teatava hinnangu sellele, milline seis valitseb veebis hetkel.

VIITED

Allsopp, J. (2005). *Are we there yet? Best practices in web development in Australia - a survey*. http://westciv.com/style_master/house/good_oil/best_practices/ ; Slaidid: <http://we05.com/resources/john-allsopp.pdf> ; Audio: <http://we05.com/podcast/mp3/we05-12-john-allsopp.mp3>

Parnas, D. (2001). *How to cope with incorrect HTML*.
<http://www.ub.uib.no/elpub/2001/h/413001/Hovedoppgave.pdf> ning
<http://elsewhat.com/thesis/>

Saarsoo, R. (2005) (1). *Standardid ja Eesti veebimaastiku olukord nende osas*.
(Proseminaritöö)
http://www.triin.net/2005/04/22/Standardid_ja_Eesti_veebimaastiku_olukord_nende_osas

Saarsoo, R. (2005) (2). *Web Standards in Estonia vol 2*.
http://www.triin.net/2005/09/18/Web_Standards_in_Estonia_vol_2