

TÜ HAAPSALU KOLLEDŽ
Infotehnoloogia osakond

Rene Saarsoo

**STANDARDID
JA
EESTI VEEBIMAASTIKU OLUKORD
NENDE OSAS**

Proseminaritöö

Juhendaja: mag. Jaagup Kippar

Haapsalu 2005

1. SISUKORD

| | |
|---|----|
| Sissejuhatus..... | 4 |
| Taust..... | 4 |
| Töö probleem ja eesmärk..... | 5 |
| 1. Materjal ja metoodika..... | 6 |
| 1.1. Valim..... | 6 |
| 1.2. Massvalideerimise programm..... | 7 |
| 1.3. Uurimuse läbiviimise aeg..... | 8 |
| 2. Tulemused..... | 9 |
| 2.1. Väljapraagitud leheküljed..... | 9 |
| 2.2. Valideerunud leheküljed..... | 10 |
| 2.3. Dokumenditüübid..... | 11 |
| 2.4. Kodeeringud..... | 13 |
| 2.5. HTML-i elemendid..... | 15 |
| 2.6. Lehed tabelitega ja ilma..... | 20 |
| 2.7. Validaatori veateated..... | 20 |
| 3. Analüüs..... | 28 |
| 3.1. Võrdlus varasemate uuringutega..... | 28 |
| 3.2. Järeldused..... | 30 |
| Kokkuvõte..... | 33 |
| Saavutatud eesmärgid..... | 33 |
| Hinnang kasutatud metoodikale..... | 33 |
| Ettepanekud uuringu jätkamiseks..... | 34 |
| Viited..... | 35 |
| Lisa 1 | |
| Kasutatud programmi lühike dokumentatsioon..... | 36 |
| Üldine..... | 36 |
| Nõudmised süsteemile..... | 36 |
| Sisend ja väljund..... | 37 |
| Kasutamine..... | 37 |
| Veateated..... | 39 |

| | |
|---|----|
| Lisa 2 | |
| Tabel kõigi W3C validaatori veateadetega..... | 41 |
| Lisa 3 | |
| Valideerunud lehekülgede loetelu..... | 44 |

Kogu uurimus koos lisadega on digitaalkujul kättesaadav aadressilt

http://triin.net/2005/04/22/Standardid_ja_Eesti_veebimaastiku_olukord_nende_osas

SISSEJUHATUS

Taust

Veebistandardid (*World Wide Web Consortium*'i ehk W3C soovitusel) on liikvel olnud juba pikemat aega – HTML 2.0 (*Hypertext Markup Language*) aastast 1995 ja CSS 1 (*Cascading Style Sheets*) aastast 1996 – sellegipoolest on vastavate standardite järgimine veebiarendajate seas äärmiselt madal.

Tabelite kasutamine puht kujunduslikel eesmärkidel oli algselt vaid kaval trikk saavutamaks välimust, mille teostamist tollane HTML ette ei nänud, ent on seniajani laialdaselt levinud. CSS – veebi kujundamiseks loodud keel – ei leidnud arendajate seas just sooja vastuvõttu, sest brauseritootjate implementatsioonid olid puudulikud ja vigased. Vahepeal võitis „brauserite sõja“ Microsoft Internet Explorer (IE) ning selle veebilehitseja senini kestva ülevõimu ajal kasvas üles hulk arendajaid, kes kirjutasid oma lehtedele „Töötab vaid IE-ga“ ning blokkisid kõik teisi brauserid kasutavad külastajad. Vahepeal õpetasid teised tootjad oma brauserid matkima IE-d – tutvustama ennast serverile valenime all ning matkima isegi IE vigu. Sedaviisi tulid ka teised brauserid areenile tagasi ning veebilojad olid sunnitud mõtlema ka nende peale. Peamised brauserid said nüüd kõik rahuldavalt hakkama CSS-iga ning tabelid võinuks vajuda unustusse, kui üksnes kogu veeb poleks kihanud õpetustest „kuidas kujundada tabelitega veebilehti“. Loomulikult olid olemas W3C soovitusel, aga neid igavaid dokumente ei lugenud ükski „terve mõistusega“ disainer. Siiski leidsid isikud, kes pidasid oma südameasjaks kuulutada sõnumit veebistandardite põhisest disainist. Raamatulettidele jõudis Jeffrey Zeldman'i lõputult ülistatud oranžide kaantega „*Designing with web standards*“, mis pööras oma lugejad lehekülghaaval veebistandardite usku. Paljud olid varmad muutuma, ent väga paljud... mitte.

Selline on taust olukorrale mis valitseb praegu maailmas ja ka Eestis. Mitmed noored ja hakkajad on pühendanud end standardite uurimisele ja järgimisele, aga väga paljud vanade võtetega harjunud pole kaugeltki varmad oma harjumuspärasest stiili muutma. Vaidlus veebistandardite heade ja halbade külgede üle ei ole aga selle teadustöö teemaks.

Töö probleem ja eesmärk

Ehkki laialt on teada, et veebistandardite järgimine pole enamiku veebiarendajate prioriteetide hulgas, pole võimalik siiski väita mitte midagi kindlat seni kuni pole teostatud tõsiseltvõetavaid uuringuid antud vallas. Nagu on pikemalt kirjeldatud selle proseminaritöö uurimusprojektis (Saarsoo 2004), on kõik teadaolevad uuringud nii eestis kui mujal maailmas olnud üsna väikesemahulised ja oluliste puudustega.

Käesoleva uuringu eesmärgiks on uurida kui võrd vastavad Eesti veebilehed W3C HTML ja XHTML (*Extensible Hypertext Markup Language*) standarditele (ehk valideeruvad). Lihtsaim vahend sellise vastavuse kontrollimiseks on kasutada tarkvaralist programmi, mis kontrollib (X)HTML dokumendi vastavust temas defineeritud dokumenditüübi deklaratsioonile. Sellistest programmidest on vaieldamatult tuntuim *W3C Markup Validation Service* (tuntud kui W3C validaator), mida käesolev uurimustöö oma eesmärgi saavutamiseks rakendada kavatsebki.

Uurimustöö peaks muu hulgas andma vastused järgmistele küsimustele:

1. Kui suur osa Eesti veebilehtedest valideerub?
2. Millised on levinuimad vead veebilehtede koodis?
3. Milline on erinevate dokumenditüüpide kasutamise osakaal?
4. Milline on erinevate kodeeringute kasutamise osakaal?

1. MATERJAL JA METOODIKA

1.1. Valim

Uurimuse jaoks valiti Neti Eesti WWW serverite nimekirjast kõik aadressid, mis olid kujuga `www.*.ee` (näiteks `www.host.ee`, `www.hk.host.ee`, `www.host.pri.ee` jne).

Sarnaselt Veiko Sangi poolt teostatud uuringuga (2004) kasutati vaid `www`-ga algavad aadresse, sest väga paljudel juhtudel on identsed need leheküljed, mille aadressidest üks algab `www`-ga ja teine mitte, näiteks `www.host.ee=host.ee`. Ühtlasi pole aadressid, mis `www`-ga ei alga, tihtipeale asutuse veebi avalikuks osaks (`intranet.host.ee`, `admin.host.ee`, `mail.host.ee`, `webmail.host.ee` jne).

Kuna eesmärgiks oli just Eesti veebilehtede uuring, siis valiti vaid `ee`-lõpulised aadressid, sest need saavad olla vaid Eesti ettevõtetel, organisatsioonidel või eraisikutel. Erinevalt näiteks Veiko Sangi serverikeskkonna uuringust, oli oluline, et valimisse satuksid just Eesti lehed (seejuures mitte tingimata eesti keelsed lehed), mitte lihtsalt eesti serverites asuvad leheküljed (`.com`, `.net`, `.org` jne), mille omanikeks võivad olla ka välismaised ettevõtted.

Valideeriti vaid esilehti, kuna nendega peaksid autorid olema kõige rohkem vaeva näinud ning – võrdlemise eesmärgil – kuna ka varasemad valideerimise uuringud on piirdunud vaid esilehtedega. Pealegi oleks nii suure hulga alamlehekülgede valideerimine osutunud ajaliselt väga mahukaks.

Uurimuses kasutati 21 905 aadressi, mis võeti Neti Eesti serverite nimekirjast (<http://www.neti.ee/cgi-bin/serverid>) seisuga 11. veebruar 2005.

1.2. Massvalideerimise programm

Kuna tuhandete lehtede ükshaaval validaatorile ettesöötmine ja tulemuste kirjapanemine on ühe inimese jaoks kaugelt liialt suur ja vaearikas töö, siis sai loodud programm, mis ilma kõrvalise sekkumiseta saaks hakkama selle massilise valideerimise protsessiga.

Nimetatud programm vaatas üksteise järel läbi kõik talle ette antud aadressid ning teostas kõigi puhul järgnevad toimingud.

Kõigepealt üritas programm lehekülge alla laadida. Kui see õnnestus, siis üritas programm veenduda, et tegemist on ikkagi veebilehega (otsides sealt üldlevinud HTML-i elemente). Lisaks kontrollis programm, et tegu poleks mõne veateate lehega (otsides `<title>` elemendist sõnu „Viga“ ja „Forbidden“), et tegu poleks suletud leheküljega (fraas „Teenus peatatud“), ning et tegu poleks leheküljega, mis suunab ümber teisele aadressile (otsides mustrit, mis oleks sarnane HTML-le:

```
„<meta http-equiv="refresh" content="5; URL=http://foo" />“.
```

HTTP päistes leiduvat infot programm ei kontrollinud.

Seejärel loendas programm 38 levinuima HTML-i elemendi esinemissagedused antud dokumendis ning üritas saata lehekülje W3C validaatorile analüüsimiseks. Kui validaator ei saanud lehte valideerida vigase kodeeringu tõttu, siis saatis programm dokumendi uuesti validaatorile ning nõudis, et too kasutaks lehekülje valideerimisel ISO-8859-1 kodeeringut (reeglina seisneb probleem selles, et leheküljel on kodeering määramata jäetud ning validaator kasutab vaikimisi UTF-8 kodeeringut, mis aga ei luba teatud baidi-kombinatsioone).

Kui programmil oli õnnestunud validaatori käest kätte saada lehekülg, millel puudusid veateated tõrgete kohta valideerimise protsessi läbiviimisel, siis salvestas programm valideerimiseks kasutatud dokumenditüübi, ning kui dokument sellele ei vastanud, loendas

kõikide veateadete esinemiste arvud.

Täpsema ülevaate programmist annab lisa 1.

1.3. Uurimuse läbiviimise aeg

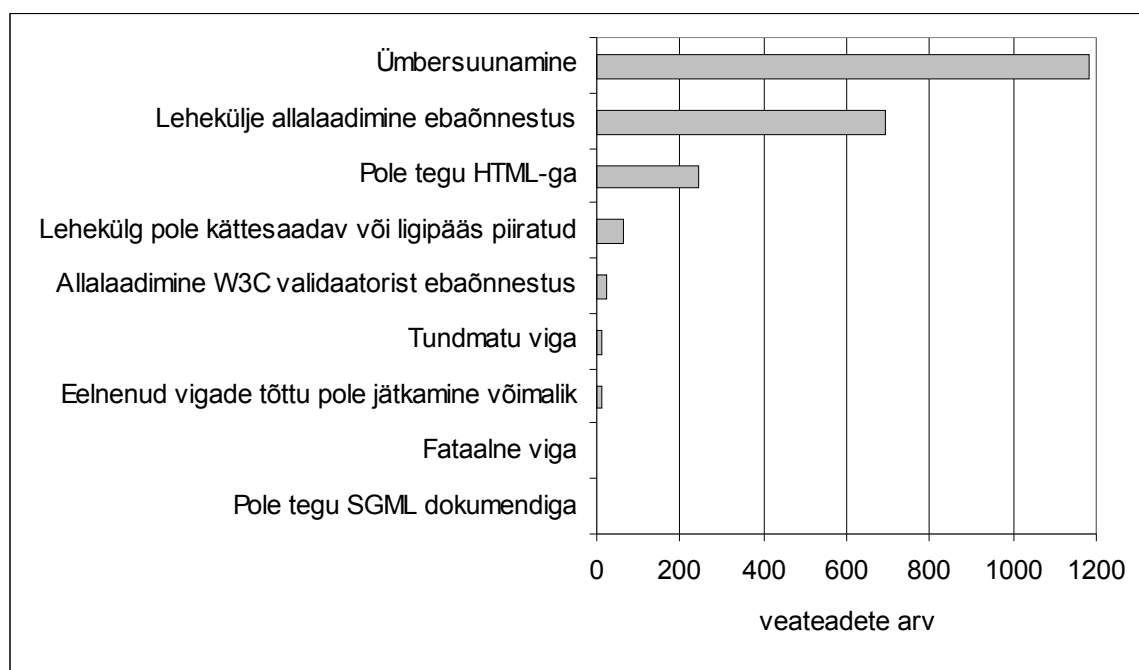
Uurimus viidi läbi ajavahemikus 2005-02-11 17:37 kuni 2005-02-13 17:29 Eesti aja järgi.

2. TULEMUSED

Käesoleva uurimuse tulemused on täies mahus kättesaadaval aadressil <http://www.triin.net/2005/04/22/tulemused.tar.gz>

2.1. Väljapraagitud leheküljed

Programm praakis töötamise käigus 21 905 aadressi seast välja 2250 (10%) ja seega jäi reaalseks valideerimiseks alles 19 655 aadressi.

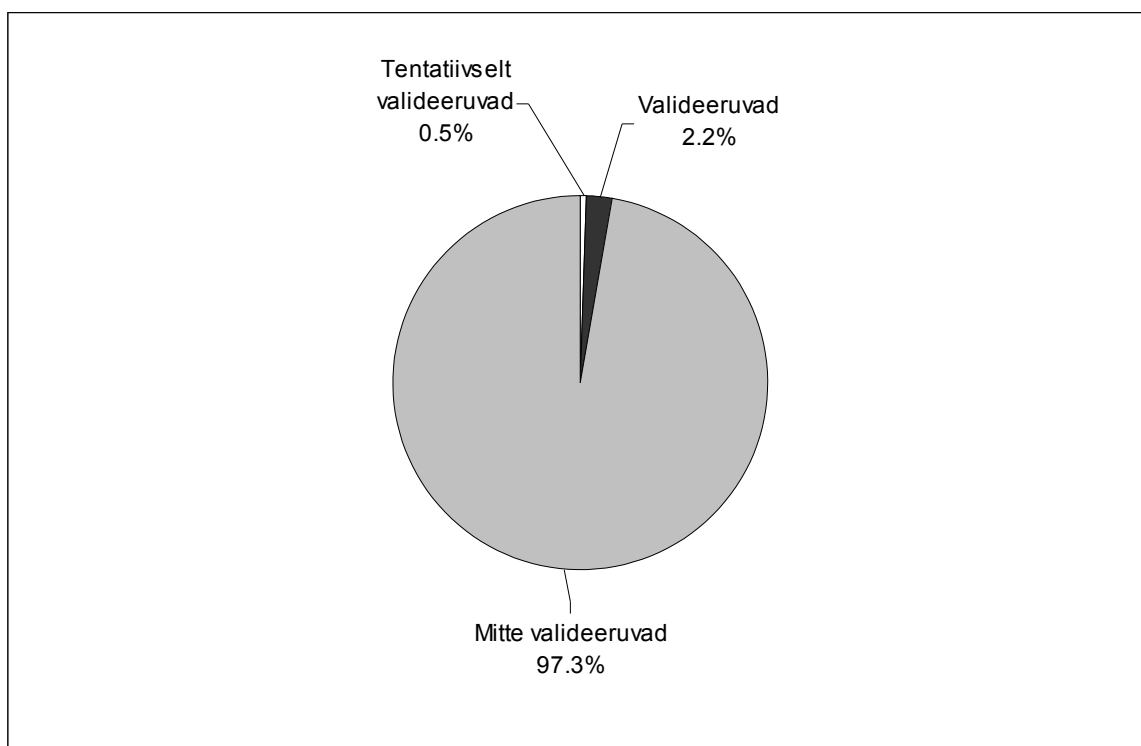


Joonis 1. Erinevate vigade esinemissagedus. (Vigade selgitused on toodud lisas 1)

Peamiseks põhjuseks lehekülgede väljapraakimisel oli ümbersuunamine teisele aadressile, teiseks ebaõnnestumine serveriga kontakteerumisel ja kolmandaks igasuguse HTML-i puudumine leheküljel (joonis 1). Täiendav info kõigi programmi poolt antavate veateadete kohta on ära toodud lisas 1.

2.2. Valideerunud leheküljed

Valideerunud lehekülgede osakaal oli, nagu eelnevalt arvata võis, väike – vaid 2,2% (436 lehekülge) valideerus täielikult, lisaks 0,5% (94 lehekülge) valideerus mööndustega (*tentatively*) (joonis 2). Sellise ääri-veeri valideerumise põhjuseks oli reeglina dokumendi kodeeringu määramata jätmine.



Joonis 2. Valideerunud lehekülgede osakaal.

42 lehte 94-st mööndustega valideerunust osutus <http://www.artfotoplus.ee/> virtuaal-

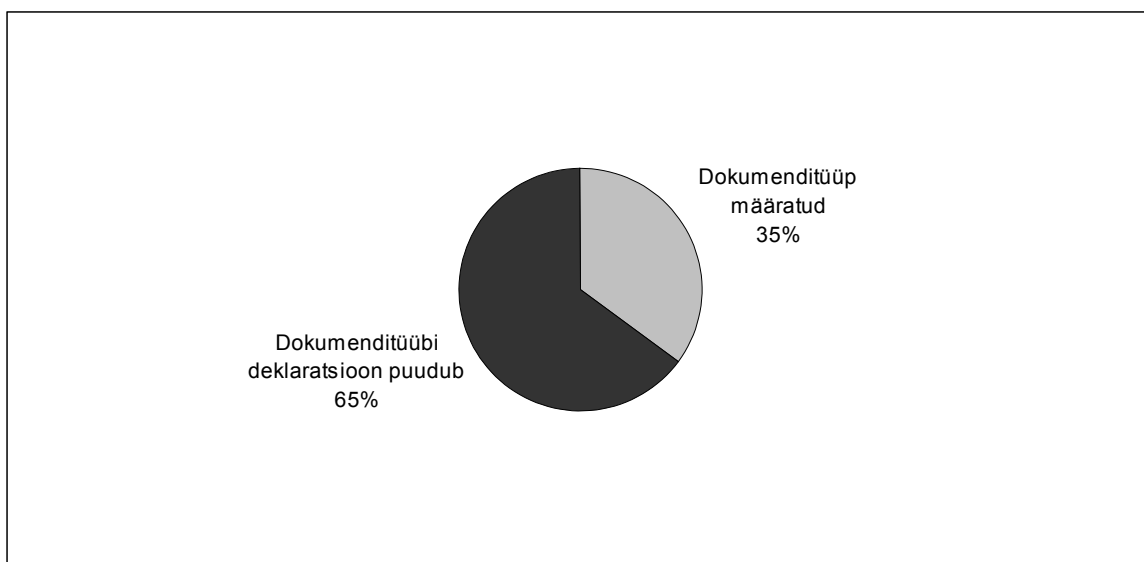
majutust kasutavateks pooleliolevateks lehekülgedeks. Ülejäänud olid kas tuvastamata jäänud ümbersuunamised, pooleliolevad lehed, veateated või lihtsalt väga vähese sisuga lehed.

NB! Kõigis järgnevates analüüsides, kus võrreldakse valideerunud ja mittevalideerunud lehekülgi, on käsitluse alt välja arvatud kõik mööndustega valideerunud lehed.

Nimekiri kõigist valideerunud lehekülgedest on ära toodud lisas 3.

2.3. Dokumenditüübid

Kõigil valideerunud lehtedel (ka mööndustega valideerunutel) oli loomulikult dokumenditüüp määratud, sest selle alusel nad valideerusidki. Mittevalideerunud lehtedel oli aga dokumenditüübi määratlus olemas vaid 35%, ehk vaid igal kolmandal leheküljel (joonis 3).



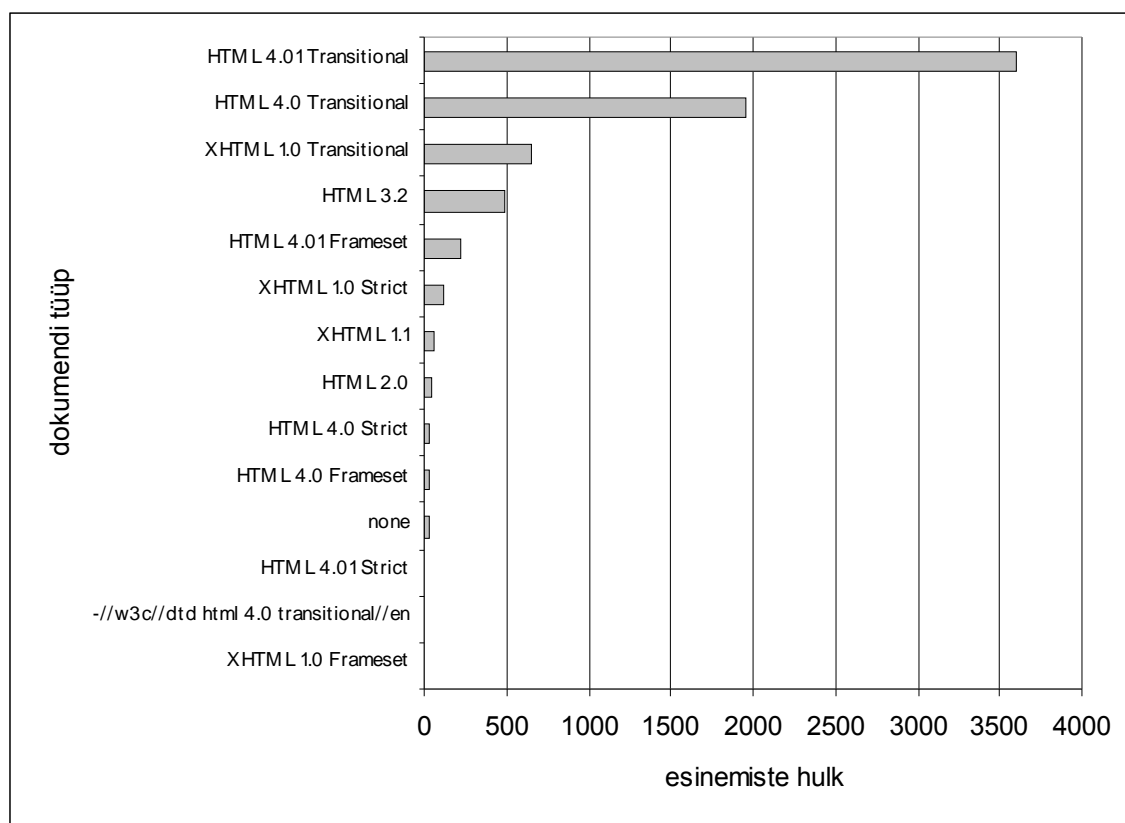
Joonis 3. Dokumenditüübi määrangu olemasolu mittevalideerunud lehtedel.

Nagu nähtub joonistelt 4 ja 5, oli levinuimaks dokumenditüübiks ülekaalukalt *HTML 4.01*

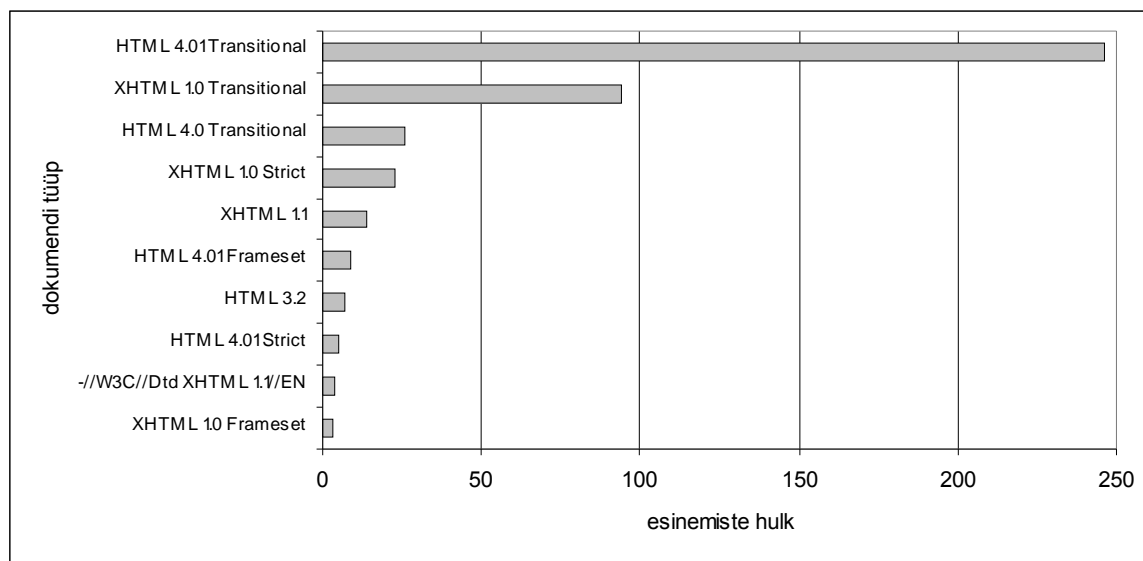
Transitional. Paljuski on selline ülekaal seletatav sellega, et enamik HTML-i redaktoreid lisab vaikumisi dokumendile just selle. Ühtlasi on tegemist ka kõige paindlikuma dokumenditüübiga, mis lubab kõiki endiselt nii levinud presentatiivseid HTML-i elemente.

Teisel kohal on *HTML 4.0 Transitional*, kuid kui vaadata vaid valideerunud lehti, siis nende seas on teisel kohal *XHTML 1.0 Transitional*, mis esimeste puhul on kolmandal kohal. Alles seejärel tuleb *XHTML 1.0 Strict*. HTML-i Strict versioon on aga vähelevinud.

Üllataval kombel on küllaltki levinud *XHTML 1.1*, mida W3C soovitusel tuleks serveerida kui `application/xhtml+xml` ent kindlasti mitte kui `text/html`. Nii palju kui vastava dokumenditüübi määranguga dokumente sai üle vaadatud kasutasid kõik ebakorrektselt ikkagi `text/html` tüüpi.



Joonis 4. Dokumenditüüpide jaotus kõigi lehtede lõikes.



Joonis 5. Dokumenditüüpide jaotus valideerunud lehtedel.

2.4. Kodeeringud

Kodeering oli HTML-is määratud vaid 68% lehekülgedest (tuleb arvestada, et osade lehtede puhul võis kodeering olla seatud HTTP päisega või XML-i deklaratsiooniga, mida kasutatav programm ei registreerinud, kuid kumbki meetod pole just kuigi levinud). Sealjuures teatas 12% lehekülgede puhul validaator probleemist kodeeringuga, mis enamikul juhtudest pidi kindlasti olema tingitud just sellest, et dokumendi kodeering oli jäetud määramata.

Järgnevas tekstis kasutatud informatsioon erinevate kodeeringute kohta pärineb peaaesjalikult Wikipedia'st (Category:Character sets).

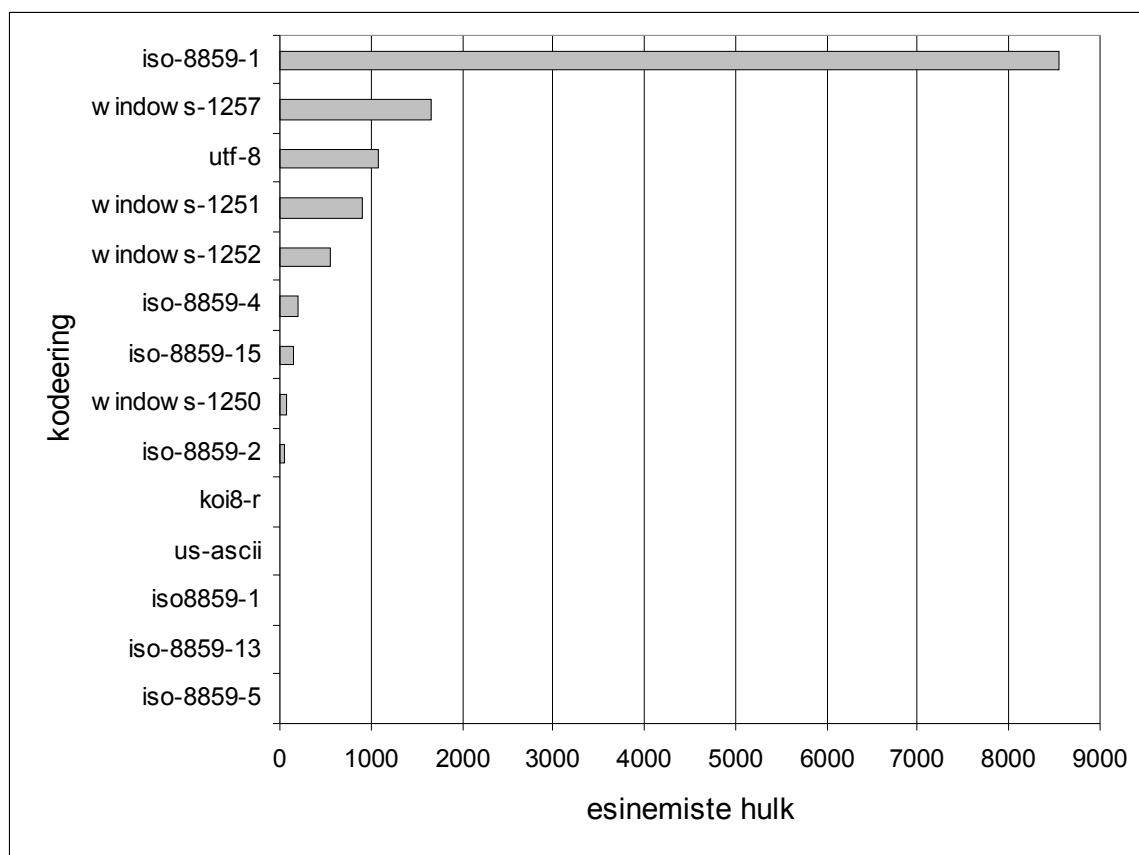
Enimkasutatavaks kodeeringuks oli ülekaalukalt ISO-8859-1 (joonis 6), mis aga ISO standardi järgi ei sobi päris hästi eestikeelsete tekstide tarbeks, sest puuduvad ž ja š. Nimetatud tähed on aga olemas Microsoft'i Windows-1252 kodeeringus, mis täiendab ISO-8859-1 kooditabelit lisasümbolitega. Teisel kohal olev Windows-1257 ja

ISO-8859-13, mida eelnimetatu täiendab, sobivad aga eesti keele jaoks kenasti. Eesti keele jaoks ametlikult soovitatav ISO-8859-15 on aga alles seitsmendal kohal, mis võib paljuski tuleneda sellest, et paljud HTML-i redaktorid toetavad vaid piiratud hulka erinevaid kodeeringuid ning ISO-8859-15 nende hulka tihtipeale ei kuulu.

Spetsiifiliselt kirillitsa tarbeks mõeldud kodeeringutest oli levinuim Windows-1251, millele järgnesid KOI8-R ja ISO-8859-5.

Universaalne UTF-8 leidis kasutamist vaid tuhatkonnal lehel. Parema ülevaate saamiseks, miks ja kus *unicode*'i kasutatakse, sai käsitsi läbi vaadatud 50 UTF-8-t kasutatavat lehekülge. Selgus, et:

- 21 nendest olid eestikeelsed,
- 13 eesti-inglise-vene-keelsed,
- 7 eesti-inglise-keelsed,
- 2 inglisekeelsed,
- 2 viie-keelsed,
- 1 eesti-vene-keelne,
- 4 ei kasutanud aga hoolimata määrangust *unicode*'i võimalusi üldse:
 - 1 kasutas vastavat kodeeringut vaid raame määraval lehel ja
 - 3 kasutasid ö asemel ö , ä asemel ä jne.



Joonis 6. Erinevate kodeeringumääratluste esinemissagedus. ISO-8859-1 on kasutusel nii ülekaalukalt, et seda on 12 korral ka ühtviisi valesti kirjutatud. (Jooniselt on välja jäetud alla kõik kodeeringunimetused, mida leidis alla kümne.)

2.5. HTML-i elemendid

Enimkasutatavaks HTML-i elemendiks oli ``, mida on suurtes kogustes kasutatud kujundamiseks. Linkide `<a>` rohkus on samas täiesti normaalne, sest ilma nendeta poleks HTML hüpertekst. Järgmised elemendid – `
`, `<table>` ja `` – ilmestavad selgelt laialtlevinud veebiehituse stiili: tabelid paigutamiseks, reavahetused lõikude eraldamiseks ja `` teksti kujundamiseks. (Joonis 7.)

Joonis 8 näitab protsentuaalselt, milliseid elemente üldse lehtedel esineb, ja illustreerib

valitsevat olukorda natuke paremini. Mõistagi kohtab pea kõigil lehtedel elemente `<html>`, `<title>`, `<head>` ja `<body>` – see, et viimast on kolme protsendi jagu vähem on seletatav sellega, et mitmed lehed, mis kasutavad raame `<frameset>`, ei vaevu kehaosa üldse lisama. `<meta>` ohtrus tuleneb vajadusest mitte-inglisekeelsetele lehtedele õige kodeering määrata.

Vaadates piltide ja tabelite esinemissagedusi võib öelda, et umbes $\frac{3}{4}$ lehekülgedest kasutavad paigutuseks tabelit (esilehele enamasti ei panda tabelit, kus oleks mingit sorti tabulaarne info nagu näiteks toodete nimekiri) ja kujunduseks pilte HTML-is (*contra* CSS-iga määratud taustapildid).

77% lehtedest kasutab kas reavahetust `
`, paragrahvi `<p>` või mõlemat. Sealjuures 40% lehtedest on nii paragrahv kui reavahetus, 21% kasutab vaid reavahetust ja 16% üksnes paragrahvi märgendit.

60% lehekülgedest leidub kas elementi `<link>`, `<style>` või mõlemaid. Kuna esimese peamine kasutusala on viitamine välimisele stiililehele, siis võib öelda, et 60% lehtedest kasutab CSS-i vähemalt mingil määral. (Küllap on ka lehti, mis rakendavad ohtralt HTML-i `style`-atribuuti, aga see on umbes samale efektiivne kui ``-i kasutamine.)

45% lehtedest kasutab teksti rõhutamiseks elemente `` ja/või ``. Nendest 63% kasutab ``-d, 24% ``-i ning 13% mõlemat korraga.

7% lehtedest kasutab teksti rõhutamiseks elemente `<i>` ja/või ``. Nendest 70% kasutab `<i>`-d, 27% ``-i ning 3% mõlemat korraga.

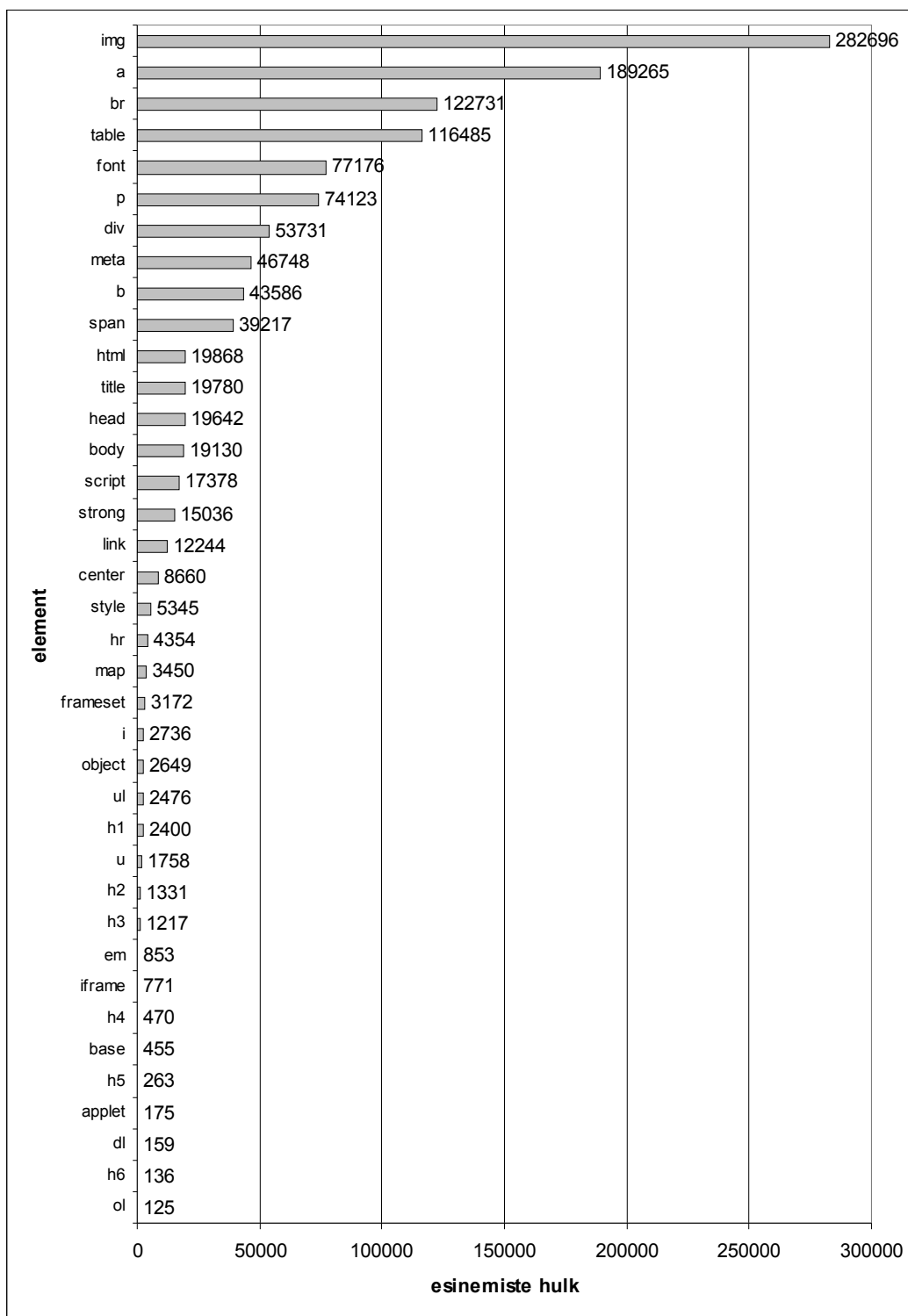
Tervelt 44% lehtedest kasutab vähemal või rohkemal määral skriptimist (element `<script>`); suures enamuses on tõenäoliselt eesmärgiks piltide vahetamine hiirega ülelibistamisel.

Palju kasutatakse mitmesuguseid iganenud vahendeid nagu 42%, <center> 21%, <map> 12%, <hr> 9% ja <u> 3%.

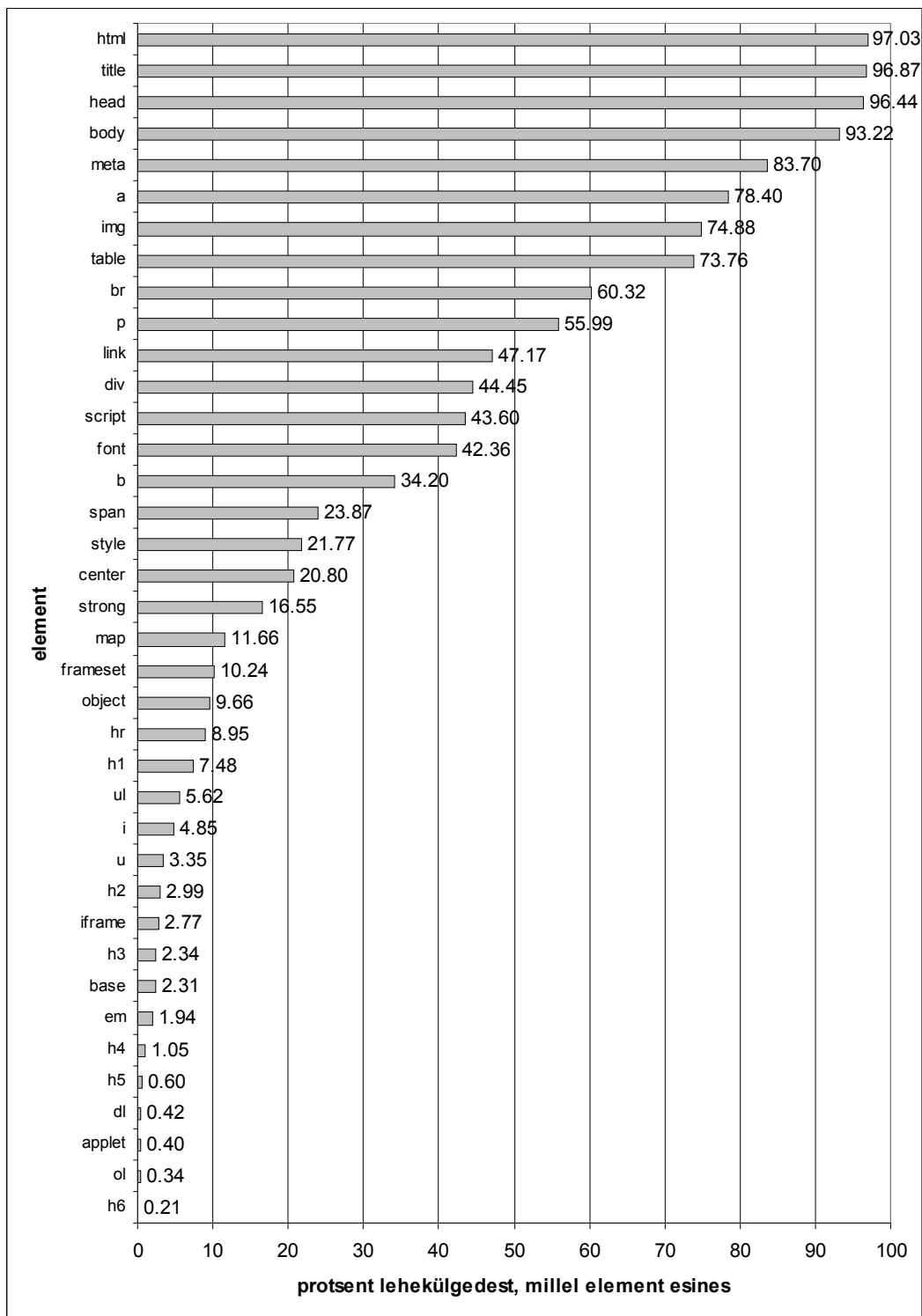
10% lehtedest kasutab raame <frameset> ning peaaegu 3% reasiseseid raame <iframe>, mis teeb kokku 13% üht või teist sorti raamidega lehti.

Pealkirjade kasutamine on kasiin – esimese taseme pealkirja <h1> kasutatakse vaid 7% lehekülgedest, <h2>-te vaid 3% ja ülejäänuid veelgi vähem.

Kuna <embed> element jäi vaatluse alt välja (kuna see puudub HTML 4.01 spetsifikatsioonis (W3C, 1999)), siis ei saa väita, et <object> on kõige populaarsem element Flash'i, Java, video, heli jms leheküljega sidumiseks, nagu joonis 8 ekslikult näitab. <object> on küll populaarsem kui <applet>, aga see viitab pigem Java vähestele kasutamisele.



Joonis 7. HTML-i elementide jaotus koguhulga põhjal.



Joonis 8. HTML-i elementide esinemissagedus.

2.6. Lehed tabelitega ja ilma

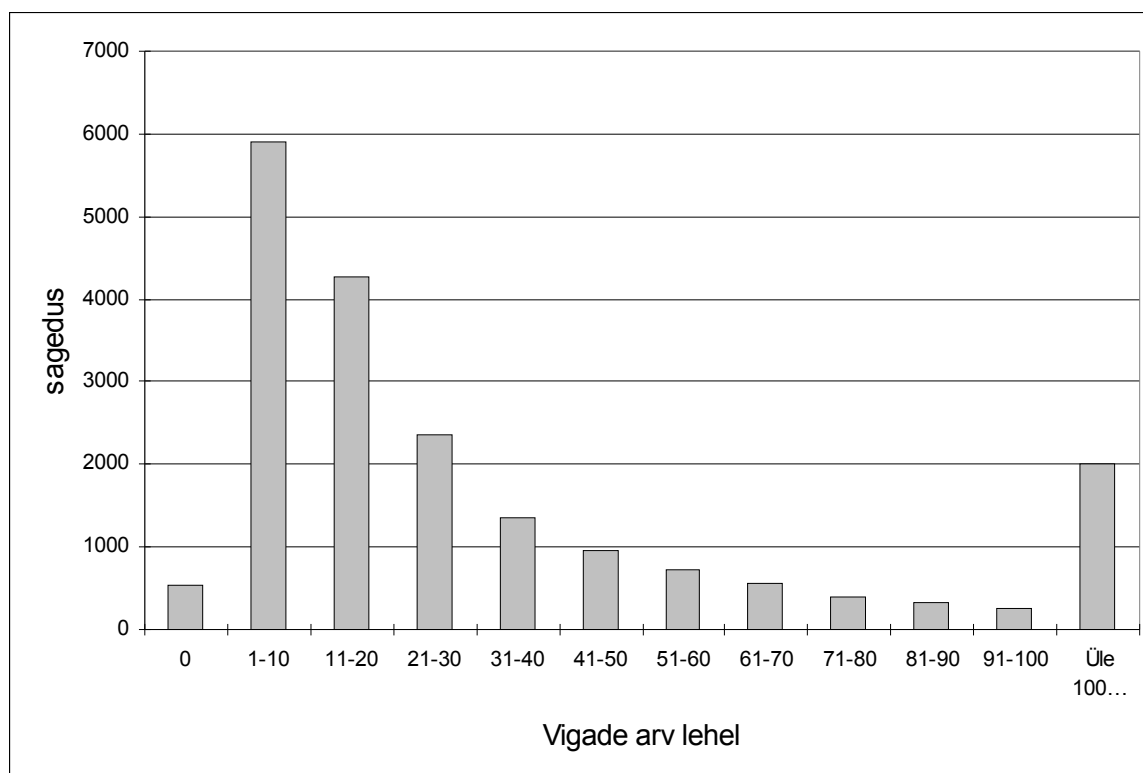
Võrreldes valideerunud ja mittevalideerunud lehekülgi oli ilma tabeliteta valmistatud lehekülgede osakaal valideerunud lehekülgede puhul pea kaks korda suurem. Samuti on valideerunud lehtede hulgas rohkem neid, kus kasutatakse vaid üht tabelit (joonis 9). Ka enim tabelleid ühe lehe kohta leidis ühel mittevalideerunud leheküljel – tervelt 238.



Joonis 9. Tabelid valideerunud lehekülgedel (vasakul) ning mittevalideerunud lehekülgedel (paremal).

2.7. Validaatori veateated

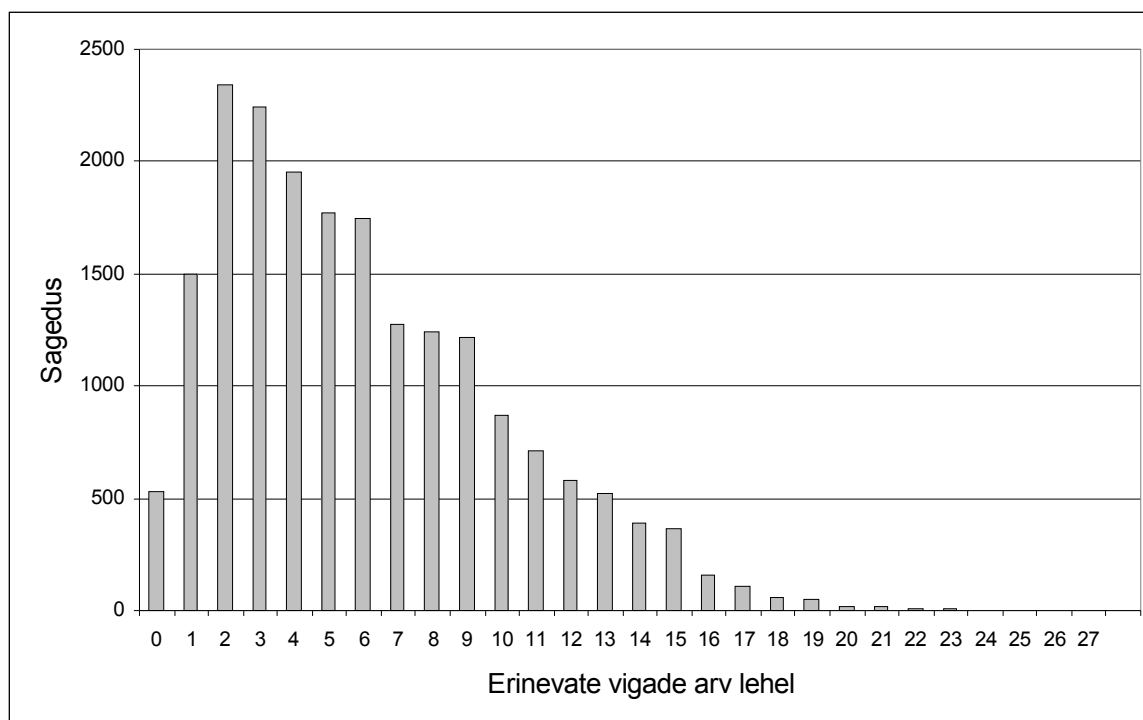
19 655-lt leheküljelt korjas programm kokku 865 602 veateadet, registreerides 88-t erinevat veateatemustrit. Keskmine vigade arv lehekülje kohta oli 44, ent seda numbrit paisutasid suureks mitmed äärmuslike vigade hulgaga lehed, nagu näiteks suurima veateadete hulgaga leht www.jogevalv.ee, mille kohta validaator andis tervelt 3177 veateadet. Sagedaseim oli 5 veateadet lehekülje kohta, mediaan oli 18. Jooniselt 10 on selgelt näha, et põhiliselt kõikus veateadete arv kusagil 1 ja 40 vea vahel.



Joonis 10. Valideerumisvigade sagedustabel.

Kuid puhtalt veateadete arvu järgi lehti hinnata võib olla üsna ülekohtune, sest kui lehekülje autor on eksinud mõne reegli vastu, siis tõenäoliselt on ta seda eksimust ka korranud, ja nii võibki olla, et kõik 1000 veateadet ühel lehel viitavad ühele ainsale veale kuid samas 5 veateadet mõnel teisel lehel annavad teada viiest erinevast probleemist.

Loendades vaid erinevaid veateateid saame hoopis teistsugused numbrid: keskmiselt 6 viga lehe kohta ning maksimaalselt 27 (nõnda palju oli erinevaid viga aadressil www.katoliku.ee). Kõige sagedamini 2 viga lehel ning mediaan 5. Erinevate vigade jaotust ilmestab ka joonisel 11 toodud sagedustabel.



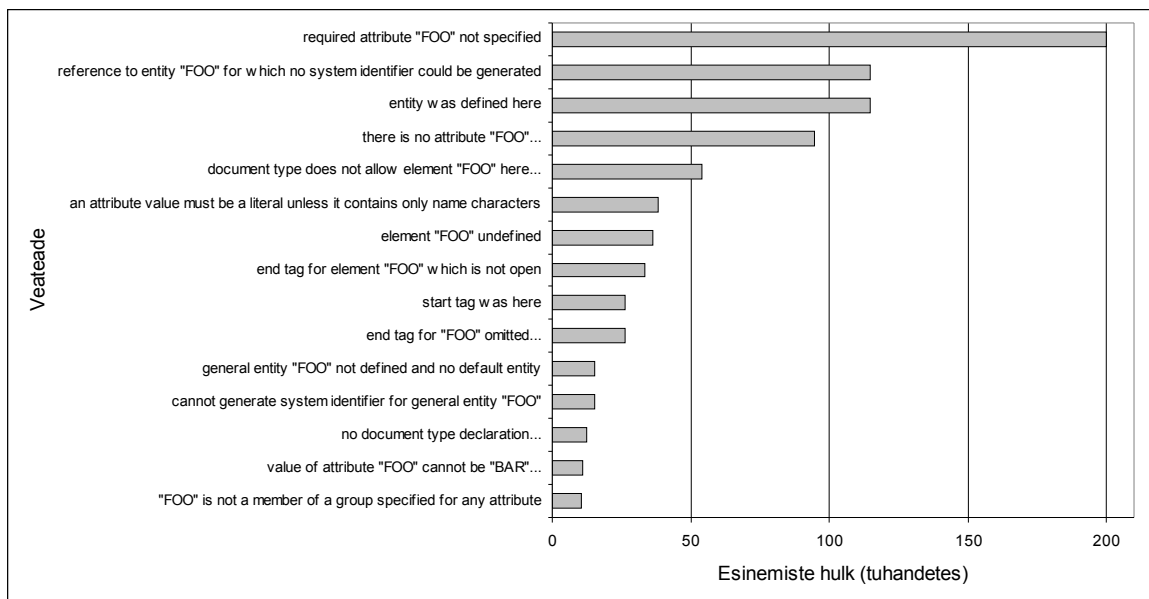
Joonis 11. Erinevate valideerumisvigade sagedustabel.

Kokkuvõttev võrdlus kõikide veateadete ja erinevate veateadete analüüsil kasutatud erinevate statistikute vahel on toodud tabelis 1.

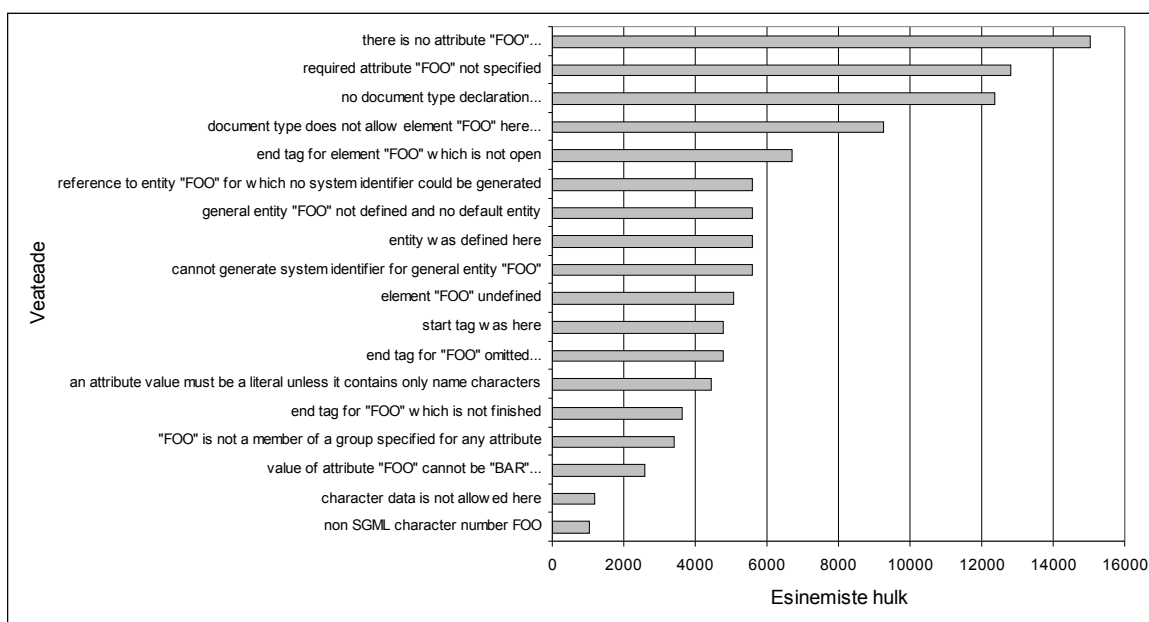
| <i>veateated</i> | <i>keskmine</i> | <i>k.keskmine</i> | <i>maksimum</i> | <i>mediaan</i> | <i>mood</i> | <i>std. hälve</i> |
|------------------|-----------------|-------------------|-----------------|----------------|-------------|-------------------|
| Kõik | 44,04 | 39,31 | 3177 | 18 | 5 | 100,04 |
| Erinevad | 6,01 | 5,97 | 27 | 5 | 2 | 4,08 |

Tabel 1. Veateadete statistikud (aritmeetiline keskmine, kaalutud keskmine, maksimaalne ja minimaalne veateadete arv, mediaan, mood ja standardhälve). Kaalutud keskmise puhul on välja arvatud 1% kõige suurema ja 1% kõige väiksema vigade arvuga lehti.

Järgnevalt vaatleme erinevate veateadete hulka (joonis 12) ja esinemissagedusi (joonis 13); arvandmed kõigi veateadete kohta on toodud ära lisas 2.



Joonis 12. Erinevate veateadete hulk. (Vaid 15 sagedasemat veateadet on näidatud.)



Joonis 13. Erinevate veateadete esinemissagedus. (Vaid 18 sagedasemat veateadet on näidatud.)

Kõige sagedamini esinenud veateade „there is no attribute "FOO"...", tulenes

mitmesuguste ebastandardsete atribuutide kasutamisest, sagedasti oli tegemist koodiga, mis nägi välja järgmine:

```
<body leftmargin="0" topmargin="0" marginwidth="0"
  marginheight="0">
<table height="100%">
<tr>
  <td background="/img/bg05.gif">
```

Ühtki neist atribuutidest pole mainitud W3C standardis, aga selleks pole ka vajadust, sest samasugune efekt on saavutatav, kui määrata vaid id-atribuut tabelilahtrile, ja kasutada järgnevat stiililehte:

```
body {margin: 0}
#lahter {background: url("/img/bg05.gif"); height:100%}
```

Üheks levinuimaks veateateks oli „required attribute "FOO" not specified“, mis tüüpiliselt tähendas, et piltidele () oli jäetud määramata atribuut alt, mis sisaldab alternatiivteksti juhuks, kui pilti pole saadaval. Samuti oli sagedasti määramata jäetud atribuut type, elementidele <script> ja <style>.

Tervelt 63% lehekülgedest oli ilma dokumenditüübi määranguta (veateade „no document type declaration...“).

Kõige rängemat tüüpi vead kuulusid veateate „document type does not allow element "FOO" here...“ alla. See veateade tähistas täiesti rumalaid vigu nagu <style> elemendi paigutamine kõikvõimalikesse kohtadesse väljaspool <head> elementi, tabelilahtrite <td> paigutamine väljapoole tabeliridasid <tr>, <form> elemendi paigutamine <table> ja <tr> vahele jms.

Kolmandikul lehekülgedest oli probleeme, mida validaator tähistas teatega „end tag for element "FOO" which is not open“. Reeglina oli tegu lihtsalt väikeste kirjavigadega, mis aga võivad erinevate brauserite tõlgenduses anda väga erinevaid tulemusi. Järgnev näide illustreerib tüüpilisi vigu, mis antud veateadet tekitasid:

```
<!-- begin paragraph -->
<p>Lorem ipsum dolor sit<br />
amet; just go to --> </b>home</b>.</p>
</body>
</tml>
```

Järgmised neli veateadet esinesid alati koos:

- „reference to entity "FOO" for which no system identifier could be generated“,
- „entity was defined here“,
- „general entity "FOO" not defined and no default entity“ ja
- „cannot generate system identifier for general entity "FOO"“.

Sagedasti esinev viga, mis seda põhjustas tulenes &-märgi kasutamisest veebiskriptidele antavate parameetrite eraldamisel. Kuid & on HTML-is erisümbol ning tähistab tähekoodi algust (näiteks ä on ä ja õ on õ), &-märgi enda jaoks on aga &. Järgnev näide demonstreerib levinud valekasutust ja korrektset varianti:

```
index.php?id=143&action=show      <!-- vigane -->
index.php?id=143&amp;action=show <!-- korrektne -->
```

Veateate „element "FOO" undefined“ põhjused langesid nelja kategooriasse:

- rudimentsete elementide nagu `<marquee>`, `<blink>` ja `<embed>` kasutamine,
- elementide `<frameset>` ja `<frame>` kasutamine ilma *(X)HTML X.X Frameset* dokumenditüübita,
- brauserispetsiifiliste laienduste kasutamine (eeskätt mitmete Microsofti vahenditega genereeritud lehekülgede puhul) ning
- kirjavead lubatud elemendinimede kirjutamisel.

Veateade „end tag for "FOO" omitted...” ja sellele alati järgnev „start tag was here” esinesid sagedasti lehtedel, mille dokumenditüübiks oli määratud XHTML, kuid mis ei järginud XML-i ranget reeglit selle kohta, et kõik elemendid `<foo>` peavad saama suletud elemendiga `</foo>` või olema üles märgitud lühendatult kui `<foo />`.

Veateade „an attribute value must be a literal unless it contains only name characters” esines nendel juhtudel kui elementide atribuutide väärtused olid jäetud jutumärkidest ümbritsemata ent väärtused ei koosnenud üksnes alfabeetilistest sümbolitest nagu järgnevas näites:

```
<p bgcolor=#F3F5EB>  
<a href=/baas/otsinq.cgi>
```

Veateade „end tag for "FOO" which is not finished” esines enamasti juhul, kui elemendid olid valesti üksteise sisse paigutatud, näiteks `<form>` asetatud otse elemendi `<table>` sisse või elemendid `<body>` ja `<html>` vahel.

Veateade „"FOO" is not a member of a group specified for any attribute” esines juhtudel, kui oli kasutatud mõnda tundmatut atribuuti, mis oli korraga nii atribuudi kui väärtuse tähenduses, nagu `noresize` ja `nocache` järgnevates näidetes:

```
<frame src="nav.html" name="nav" noresize>
```

```

```

Veateade „value of attribute "FOO" cannot be "BAR"..." esines reeglina juhul, kui align atribuudile oli määratud mõni muu väärtus kui top, middle, bottom, left või right nagu järgmises näites:

```

```

```

```

3. ANALÜÜS

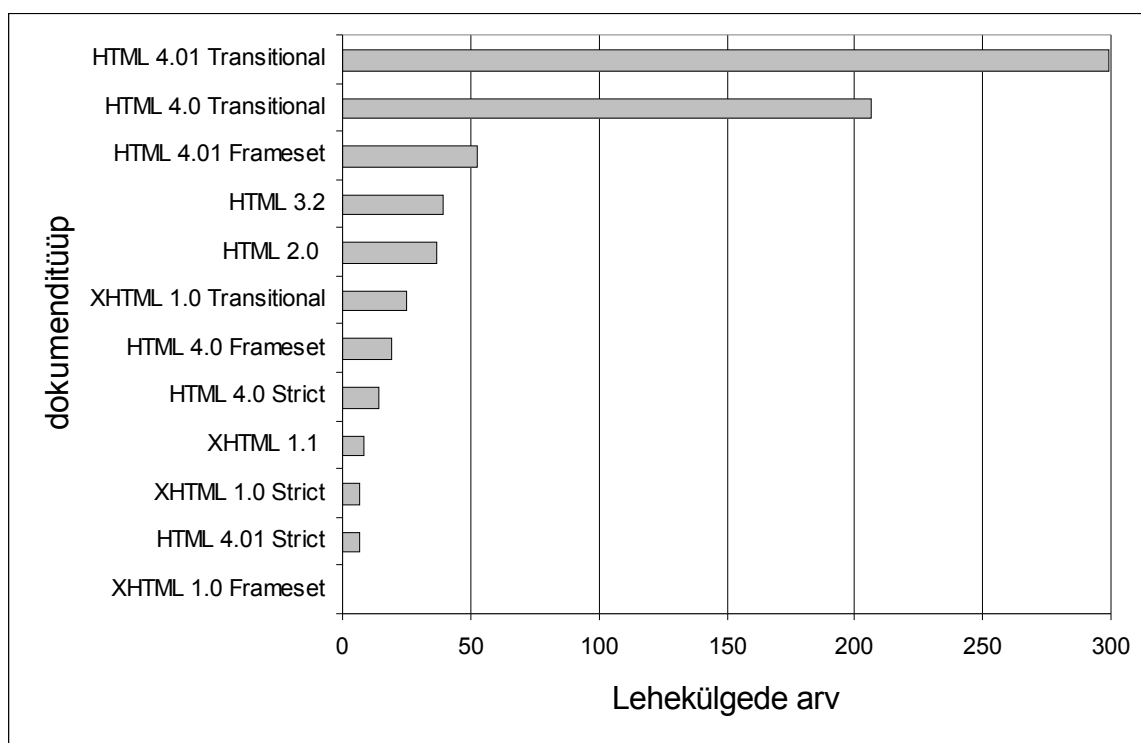
3.1. Võrdlus varasemate uuringutega

Kuna Thomas Dowling'i 1997. aasta uuringu andmed on liialt vananenud ning Greg Lanier'i 2003. aasta uuringus käsitleti vaid vähest valitud hulka lehekülgi, siis piirdub järgnev võrdlus vaid Johannessen'i 2004. aasta uuringutest pärit andmetega, kui kõige sarnasematega:

1. Soren Johannessen, „*The use of W3C standards in Denmark*“, 2004-01-31 kuni 2004-02-16, vaadeldi 2033 Taani riigiasutuse lehekülge, kes olid registreerunud konkursile „*Top of the Web*“ aastal 2003.
2. Soren Johannessen, „*Danish companies and the use of W3C standards*“, 2004-03-11 kuni 2004-03-16, vaadeldi 250 Taani ettevõtte lehekülge.
3. Soren Johannessen, „*The use of W3C standards in Denmark II*“, 2004-11-18 kuni 2004-11-28, vaadeldi 580 Taani riigiasutuse lehekülge, kes olid registreerunud konkursile „*Top of the Web*“ aastal 2004; väiksem lehtede arv tulenes sellest, et seekord oli kehtestatud osavõtumaks (Johannessen kirjavahetus 19. Aprill 2005).

Esimeses uurimuses valideerus 3,05% lehtedest, teises 0,4% (vaid üks leht 250-st) ja kolmandas 12,24%. Arvestades, et nimetatud uuringute valim oli märksa kitsam, siis pole tulemus 2,2% (koos mööndustega valideerunud lehtedega 2,7%) just väga paha (võrreldes 3,05%-ga; teine ja kolmas uurimus on võrdluseks liialt väiksed). Kuid tuleb arvestada, et mitmed käesolevas uurimuses valideerunud lehtedest olid väga lihtsakoelised, ning seetõttu pole võimalik öelda kas Eestis olukord on võrreldes Taaniga parem või halvem; kindel on vaid see, et Taani valitsus on erinevalt Eesti valitsusest tugevalt soovitanud riigiasutuste veebilehtede loomisel järgida W3C standardeid (Johannessen 2004b).

Esimeses uurimuses oli dokumentitüübi deklaratsioon olemas 35% lehekülgedest, mis langeb täpselt kokku käesoleva uurimusega. Teises Taani uurimuses oli dokumentitüübi deklaratsioon 38%, kolmandas aga tervelt 61% lehtedest.



Joonis 14. Dokumentitüüpide jaotus Taani valitsusasutuste lehekülgedel 2004. aasta jaanuari-veebruari uuringus. (Johannessen'i andmete põhjal.)

Sarnaselt käesoleva ja Taani esimese uurimusega (jättes järgnevalt vaatluse alt välja Taani väiksemamahulised uuringud) oli levinuimaks dokumentitüübiks *HTML 4.01 Transitional* (Joonis 14), millele järgnes *HTML 4.0 Transitional*. Erinevalt Eestist oli Taani lehtede puhul kolmandal kohal *HTML 4.01 Frameset*, mis siin oli alles viiendal kohal ja viitab sellele, et Eestis kasutatakse raame vähem (ent see võib olla ka ekslik, sest valimid on siiski küllaltki erinevad). raamide vähesema kasutamise põhjuseks võib olla suurem serveripoolsete skriptide osakaal. Eestis kolmandal kohal olev *XHTML 1.0 Transitional* on Taani puhul oli alles kuuendal kohal, mis näikse viitavat siinsele suuremale XHTML-i kasutamisele. Nii Eestis kui Taanis oli neljandal kohal HTML 3.2. Taani uurimuses viies

HTML 2.0 oli aga Eestis kaheksas, lubades taas oletada, et dokumenditüüpide valiku osas on Eesti progressiivsem.

3.2. Järeldused

Väga väheste lehtede vastavus veebistandarditele näib olevat tingitud kolmest peamisest standardeid pärssivast tegurist:

- üldlevinud halvad praktikad,
- teadmatus ja
- hoolimatus.

3.2.1. Üldlevinud halvad praktikad

Enamasti järgneme me nendele, kes lähevad kindla sammuga ees – kas antud rada ka õige on, on rohkem eesminejate mure – meie valime lihtsalt kõndimiseks kindlaid ja sissetallatud radu. Sama allegooria kehtib ka veebidisainerite puhul.

Kõige mõjuvõimsamaks praktikaks *mainstream* veebidisainis on tabelite kasutamine paigutuseks ning piltide `` lisamine otse HTML-i kujunduseesmärgil, mida teeb vähemalt 73% lehtedest. Endiselt on laialt levinud mitmesugune presentatiivne HTML: paljukirutud `` leidub 42% lehtedest ning elementi `<center>` kasutab 21%.

44% lehtedest kasutab JavaScript'i, mis enamiku lehtede puhul on selge liialdus – seda kasutatakse peamiselt nn. hover-efektide saavutamiseks, mida aga märksa lihtsam on teostada CSS-ga. Kuid CSS leiab kasutust vaid 60% lehtedest ja nendelgi enamasti väga vähesel määral.

3.2.2. Teadmatus

Enamik veebiautoreid ei lisa oma lehekülgedele dokumentitüübi definitsiooni, pidades seda kas mittevajalikuks või teadmata üldse sellise konstruktsiooni olemasolust. Keskmine veebidisainer lihtsalt ei näe vähimatki vajadust dokumentitüüpi määrata – pealegi ei räägi enamik HTML-i õpetusi sellest sõnagi. Need kes määravad, kasutavad enamasti üht kolmest *Transitional*-tüüpi definitsioonist, kas siis seetõttu, et need on kõige lõdvemate reeglitega, või lisab kasutatav tarkvara vaikimisi vastava definitsiooni igale dokumendile.

Eesti keeles leiduvate täpitähtede tõttu peaks suur enamus lehtedest olema sunnitud määrama kodeeringu, ometi on seda teinud vaid kaks kolmandikku. Seejuures nimetab ülekaalukas enamus lehtedest oma kodeeringuks ISO-8859-1, mis aga vastava ISO standardi järgi tegelikult eesti keele nõuetele ei vasta. Kuid kuna brauserid kasutavad nimetatud kodeeringumääratlusega dokumendi kuvamisel hoopis Windows-1252 kodeeringut, ning sama kasutavad ka paljud HTML-i redaktorid, siis pole midagi, mis nimetatud veast praktikas märku annaks.

Tuleb tõdeda, et paljud tegelikult ei tunne HTML-i või peavad presentatiivseid elemente palju väärtuslikumaks kui struktuurseid, näiteks pealkirju – suurepäraseid lehekülje struktureerimise vahendeid ning otsingumootorite „magneteid“ – kasutab vaid alla 10% lehtedest. Erinevaid loendeid kasutab alles iga kahekümnes.

Ning kui ei tunta HTML-i, siis pole imestada, et XHTML-i reeglitest teatakse veelgi vähem ning XHTML-dokumendi sisse kirjutatakse „vana hea“ HTML-i reeglite järgi, mis põhjustas mitmeid veateateid valideerimisel.

3.2.3. Hoolimatus

Üheskoos dünaamiliste veebirakenduste kasvuga on autorid oma lehe koodi suhtes veelgi

hooletumaks muutunud. Tihti minnakse lihtsaima vastupanu teed ning varustatakse reavahetused lihtsalt elemendiga `
`, PHP-s on selle tarbeks koguni sisseehitatud funktsioon `nl2br`, mille kasutust tuleks piirata arendusfaasiga ning valmisrakendustes tarvitada midagi paremat. Dünaamiliste lehtede loojad ei paista teadvat või hoolivat sellest, et `&` on HTML-is eritähendusega sümbol ning selle vea all kannatas tervelt 30% lehekülgedest (see number näib suuremana kui arvestada, et enamik lehti saavad hakkama ilma linkideta, millel oleks üle ühe parameetri).

Üllatavalt sagedasti põhjustas vigu elementide paigutamine valesse kohta, mis on selge märk hoolimatusest, sest on raske uskuda, et inimesed, kes valmistavad lehekülgi raha eest pole teadlikud sellest, et elemendi `<html>` sees saavad vahetult olla vaid `<head>` ja `<body>`.

KOKKUVÕTE

Saavutatud eesmärgid

Käesoleva uurimuse võib lugeda kordaläinuks, sest saavutati kõik eesmärgid, mis said töö alustamisel püstitatud. Lisaks koguti hulk informatsiooni, mis on heaks võrdlusmaterjaliks tulevikus tehtavatele uuringutele. Pealegi on nüüd olemas programm, mis suurema osa taolise uurimuse rutiinsest vaevast enda kanda võib võtta.

Vähene standardite järgimine ning halvad HTML-i kirjutamise praktikad on probleemid, mida võib teadvustada endale igaüks, kes vähegi vaevub populaarsete veebilehtede koodipoolt kaema, kuid seni polnud võimalik välja tuua arve, mis kirjeldaksid kui tänane olukord siiski on. Just nimelt see on lünk, mille täitmine oli käesoleva uurimuse suurim saavutus.

Hinnang kasutatud metoodikale

Mõistagi ei saa teostatud uurimust pidada ilmeksimatuks absoluutseks tõeks – kaugel sellest. Kasutatud programm ei suutnud välja praakida sugugi kõiki ümbersuunamisi, veateateid ja pooleliolevaid lehekülgi. Kuid pole põhjust arvata, et need leheküljed on uuringu tulemusi oluliselt ühes või teises suunas kallutanud. Pigem jagunesid antud lehed hajusalt valideerunud ja mittevalideerunud lehtede vahel, mõõndustega valideerunud lehti (mille hulka paljud sellised sattusid) aga võrdlustesse eelnimetatud hulkade vahel ei arvatud.

Sellegipoolest ei saa jätta tähelepanuta vajadust edasistes uuringutes metoodikat täiustada.

Ettepanekud uuringu jätkamiseks

Kõige olulisem on täiustada teatud lehtede väljapraakimise metoodikat:

- lisaks HTML-ile sisule tuleks kontrollida ka HTTP päiseinfot,
- välja tuleks arvata lehed, mis on liiga väikese mahuga (näiteks alla 500 baidi),
- probleemseid lehti tähistavate fraaside hulka tuleks oluliselt täiendada.

Nüüd, kui on teada sagedasemad veateadete mustrid, on võimalik programm õpetada ära tundma mitmeid konkreetseid tüüpprobleeme, eristamaks näiteks seda, kas veateates „required attribute "FOO" not specified“ on FOO väärtuseks `alt` või `type` või hoopis midagi muud. Ka HTML-ist otsitavate elementide hulk tuleks laiendada (kasvõi juba selleks, et uurida kui palju kasutatakse elementi `<embed>`, mida käesolev uurimus üldse ei arvestanud).

Veelgi rohkem infot annaks aga see, kui võtta uurimuse alla ka mitmed leheküljega seotud failid:

- stiililehti ning uudisvooge võiks samuti valideerida, kasutades olemasolevaid validaatoreid,
- stiililehtede ja skriptide koodi võiks analüüsida analoogselt HTML-ga ning
- uurimusse võiks kaasata veel teisi lehekülgi peale avalehe.

Antud uuringu jätkamine ei tohiks olla küsimus – see on põhimõtteliselt vaid programmi käivitamise vaev.

VIITED

Category:Character sets. *Wikipedia*. http://en.wikipedia.org/wiki/Category:Character_sets

Dowling, T. (1997). *Validating HTML*. <http://silver.ohiolink.edu/tdowling/validation.html>

Eesti WWW Serverid. Elion. <http://www.neti.ee/cgi-bin/serverid>

HTML 4.01 Specification. (1999). W3C. <http://www.w3.org/TR/html4/>

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>

Johannessen, S. (2004). *The use of W3C standards in Denmark II*.

<http://www.ae35-unit.dk/standard/bestonthenet2004.html>

Lanier, G. (2003). *Universal web design: a survey of web accessibility and usability*.

<http://wwwx.cs.unc.edu/Research/assist/et/projects/universal/>

Sang, V. (2004). *Eesti veebi serveripoolse keskkonna uuring*. [Magistritöö]. Tartu: Tartu

Ülikool. <http://math.ut.ee/~veikos/magtoo/>

W3C Markup Validation Service. W3C. <http://validator.w3.org>

LISA 1

KASUTATUD PROGRAMMI LÜHIKE DOKUMENTATSIOON

Üldine

Uurimuse läbiviimiseks loodud programm on avalikult kättesaadav aadressil <http://www.triin.net/2005/04/22/batchval.tar.gz>

Programm valideerib W3C validaatorit kasutades talle ette antud veebilehed ning salvestab statistika nii failide, validaatori veateadete kui programmi töö käigus tekkinud vigade kohta.

Terviklik programm koosneb ühest peaprogrammist ja kahest alamprogrammist. Peaprogramm on kirjutatud Python'is ning alamprogrammid Perl'is. Peaprogramm on teostatud Python'is, sest see keel võimaldab internetist faile alla laadida ja teisi programme käivitada ning nende antavaid veateateid registreerida – saavutades seda kõike väga vähesel koodihulgaga. Alamprogrammid tegelevad allalaetud failidest mustrite otsimisega ning on seetõttu teostatud Perl'is – keeles, mis on kuulus oma mitmekülgsel regulaaravaldiste poolest.

Nõudmised süsteemile

- Python (testitud versiooniga 2.3.3),
- Perl (testitud versiooniga 5.8.3),
- Linux või mõni muu UNIX-i laadne süsteem (testitud SuSE ja Gentoo Linux peal).

Sisend ja väljund

Programmi sisendiks on fail aadressidega, kus aadressid on eraldatud reavahetustega ning ei sisalda protokollitunnust, s.t. mitte `http://www.domeen.ee` vaid `www.domeen.ee`. Vaikimisi eeldatakse, et kõik aadressid kasutavad HTTP protokollit. (Vaikimisi on selleks failiks „servers.list“.)

Programmi väljundiks on kolm faili:

- veateadete fail, kuhu salvestatakse kõik programmi töö käigus tekkinud probleemid (vaikimisi „err.log“),
- lehekülje statistika fail, kuhu salvestatakse info lehelt leitud HTML-i kohta tabelina eraldades väljad komadega (vaikimisi „sitestat.csv“), genereeritud alamprogrammi „checkpage.pl“ poolt,
- validaatori statistika fail, kuhu salvestatakse info selle kohta kas leht valideerus või ei valideerunud ning statistika erinevate veateadete arvu kohta tabelina eraldades väljad komadega (vaikimisi „results.csv“), genereeritud alamprogrammi „validate.pl“ poolt.

Kasutamine

Programmiga töö alustamiseks tuleks luua uus kataloog ning sinna sisse programm arhiivist lahti pakkida:

```
$ tar xvfz bachval.tar.gz
```

Kui Perl ja Python ei asu kataloogis `/usr/bin`, siis tuleks modifitseerida failide „batchval.py“, „validate.pl“ ja „checkpage.pl“ esimesi ridu, näiteks asendada rida

```
#!/usr/bin/python
```

reaga

```
#!/usr/local/bin/python
```

Lihtsaimal juhul toimub programmi kasutamine järgmiselt (oluline on olla programmi kataloogis):

```
$ ./batchval.py --prepare
```

Programm võtab sellisel juhul oma sisendi vaikimisi kasutatavast failist ning kirjutab väljundid samuti vaikimisi märgitud failidesse. Parameeter „--prepare“ (lühendatult „-p“) käsib väljundiks kasutatavad failid eelnevalt tühjendada/luua ning paigutab nendes ühtlasi komadega eraldatud tabeli tulpade pealkirjad. Kui sooviks on näiteks poolelijäänud programmi tööd jätkata nii, et väljundifaile ei tühjendataks ja lihtsalt jätkataks nendes ridade lisamist, siis tuleb „--prepare“ parameeter ära jätta.

Soovi korral võib nii sisend kui väljundfailide nimed programmile eraldi ette määrata:

```
$ ./batchval.py --input <sisendfail>  
                --error <veateadete fail>  
                --check <HTML-i statistika fail>  
                --validation <validaatori statistika fail>
```

Vaikimisi salvestab programm allalaetud leheküljed ajutiselt faili „check.html“, sedagi faili võib ette määrata kasutades parameetrit --download <failinimi>.

Kasutades parameetrit --debug katkestab programm töö alati kui kohtab mõnda programmi jaoks tundmatut W3C validaatori veateadet. See parameeter on kasulik

täiendamaks veateatemustrite loetelu, mida sisaldab fail „validate.pl“.

Abi erinevate parameetrite kohta on alati kättesaadav käsurealt, kirjutades:

```
$ ./batchval.py --help
```

Veateated

Olulist rolli selle programmi juures mängivad veateated, mis kannavad mitmesugust olulist infot ning ei tähenda reeglina viga programmis endas vaid pigem probleeme valideerimiseks määratud lehekülgedega. Mugavamaks analüüsiks salvestatakse kõik veateated vaikimisi faili „err.log“.

Järgnevad veateated on kõige tavalisemad ning tähistavad probleeme valideeritava lehekülje või seda vahendava serveriga:

- **Page redirected** – lehekülje koodis oli ümbersuunamist tähistav <meta> element.
- **URL unreachable** – nimetatud aadressilt ei õnnestunud dokumenti alla laadida.
- **Not an HTML-document** – lehekülje koodis ei leitud ühtegi üldtuntud HTML-i elementi.
- **Page unavailable or access denied for W3C validator** – W3C validaatoril ei õnnestunud leheküljele ligi pääseda.
- **Error downloading from W3C validator** – ebaõnnestus valideerimise tulemuste allalaadimine.

W3C validaatori poolt antavad veateated:

- **Cannot continue because of previous errors** – validaator katkestab lehekülje valideerimise eelnenud vigade tõttu.

- **This is not an SGML document** – validaator ei tunnista antud faili SGML-iks (seega ei saa antud dokument olla ka ei HTML ega XHTML).
- **A Fatal Error occurred (with encoding or DTD)** – hoolimata sellest, et validaatorile määrati eraldi kasutamiseks kodeering ISO-8859-1, polnud valideerimise teostamine võimalik.

Programmi enda puudustest tingitud veateated:

- **Unknown Error message recovered** – W3C validaator andis veateate, mida programm ei tunne.
- **Error in script** – tundmatu viga alamprogrammis „validate.pl“.
- **A Mystery Error** – seni avastamata viga, mis registreeritakse failis „validate.pl“ real 676.

LISA 2

TABEL KÕIGI W3C VALIDAATORI VEATEADETEGA

Järgnev tabel kajastab kõiki validaatori poolt antud veateateid. Esimene tulp näitab kõigi registreeritud veateadete arvu, teine lehekülgede arvu, millel veateade esines. Sorteeritud viimase tulba järgi.

| Veateade | koguhulk | esinemisi |
|--|-----------------|------------------|
| there is no attribute "FOO"... | 94600 | 15034 |
| required attribute "FOO" not specified | 199807 | 12829 |
| no document type declaration... | 12375 | 12375 |
| document type does not allow element "FOO" here... | 54242 | 9241 |
| end tag for element "FOO" which is not open | 33335 | 6695 |
| cannot generate system identifier for general entity "FOO" | 15382 | 5603 |
| entity was defined here | 114835 | 5603 |
| general entity "FOO" not defined and no default entity | 15382 | 5603 |
| reference to entity "FOO" for which no system identifier could be generated | 114835 | 5603 |
| element "FOO" undefined | 36574 | 5073 |
| end tag for "FOO" omitted... | 26100 | 4764 |
| start tag was here | 26100 | 4764 |
| an attribute value must be a literal unless it contains only name characters | 38234 | 4446 |
| end tag for "FOO" which is not finished | 5366 | 3648 |
| "FOO" is not a member of a group specified for any attribute | 10380 | 3396 |
| value of attribute "FOO" cannot be "BAR"... | 10866 | 2583 |
| character data is not allowed here | 3537 | 1200 |
| non SGML character number FOO | 5763 | 1038 |
| "FOO" not finished but containing element ended | 1255 | 793 |
| duplicate specification of attribute "FOO" | 2092 | 663 |
| an attribute specification must start with a name or name token | 1501 | 511 |
| comment [declaration] started here | 1588 | 479 |
| start tag for "FOO" omitted but its declaration does not permit this | 1046 | 472 |
| invalid comment declaration... | 1554 | 469 |
| ID "FOO" already defined | 3484 | 466 |
| ID "FOO" first defined here | 3484 | 466 |
| value of attribute "FOO" must be a single token | 610 | 457 |
| syntax of attribute value does not conform to declared value | 812 | 322 |
| reference not terminated by REFC delimiter | 5228 | 286 |

| | | |
|--|------|-----|
| an attribute value specification must be an attribute value literal unless SHORTTAG YES is specified | 6786 | 275 |
| character "FOO" not allowed in attribute specification list | 640 | 268 |
| reference to external entity in attribute value | 4674 | 264 |
| "FOO" is not a reserved name | 4096 | 256 |
| invalid formal public identifier "FOO"... | 249 | 249 |
| reference to non-SGML character | 2236 | 244 |
| value of attribute "FOO" invalid | 1380 | 225 |
| "FOO" declaration not allowed in instance | 230 | 194 |
| character "FOO" is not allowed in the value of attribute "BAR" | 381 | 142 |
| an attribute value literal can occur in an attribute specification list only after a ... delimiter | 206 | 138 |
| delimiter "FOO" invalid... | 268 | 136 |
| "FOO" not finished but document ended | 140 | 112 |
| an attribute value specification must start with a literal or a name character | 292 | 105 |
| literal is missing closing delimiter | 138 | 90 |
| character "FOO" is the first character of a delimiter but occurred as data | 244 | 83 |
| character "FOO" not allowed in prolog | 107 | 67 |
| NET-enabling start-tag not immediately followed by null end-tag | 176 | 50 |
| the name and VI delimiter can be omitted from an attribute specification only if SHORTTAG YES is specified | 113 | 34 |
| unterminated comment... | 34 | 34 |
| name start character invalid... | 98 | 33 |
| unknown declaration type "FOO" | 64 | 29 |
| S separator in comment declaration | 52 | 18 |
| DTD did not contain element declaration for document type name | 17 | 17 |
| cannot generate system identifier for public text "FOO" | 15 | 15 |
| unclosed end-tag requires SHORTTAG YES | 14 | 14 |
| character "FOO" not allowed in end tag | 14 | 13 |
| entity reference illegal after document element | 19 | 13 |
| end of document in prolog | 12 | 12 |
| multiple comments in comment declaration | 2331 | 12 |
| countunclosed start-tag requires SHORTTAG YES | 22 | 10 |
| entity end not allowed in processing instruction | 10 | 10 |
| reference to non-existent ID "FOO" | 54 | 7 |
| no system id specified | 6 | 6 |
| number of open elements exceeds TAGLVL (100) | 43 | 6 |
| an entity end in a literal must terminate an entity referenced in the same literal | 5 | 5 |
| no start tag specified for implied empty element "FOO" | 4 | 4 |
| value of fixed attribute "FOO" not equal to default | 4 | 4 |
| name character invalid... | 6 | 3 |
| "FOO" is not a function name | 20 | 2 |
| entity end not allowed in attribute specification list except in attribute value literal | 2 | 2 |

| | | |
|---|----|---|
| no document element | 2 | 2 |
| "FOO" is not a character number in the document character set | 5 | 1 |
| character "FOO" invalid... | 2 | 1 |
| could not get "FOO" from "BAR"... | 1 | 1 |
| digit invalid... | 1 | 1 |
| document type "FOO" already defined | 1 | 1 |
| DTDs other than base allowed only if CONCUR YES or EXPLICIT YES | 1 | 1 |
| entity end not allowed in end tag | 1 | 1 |
| named character reference | 19 | 1 |

LISA 3 VALIDEERUNUD LEHEKÜLGEDE LOETELU

Järgnev loetelu on mugavamaks kasutamiseks kättesaadaval veebilehena aadressil

http://www.triin.net/2005/04/22/Valideerunud_lehtede_loetelu

- | | | |
|----------------------------|-----------------------------|---------------------------|
| 1. www.2web.ee | 42. www.avancia.ee | 83. www.diip.ee |
| 2. www.3paksu.ee | 43. www.away.ee | 84. www.dnbarena.ee |
| 3. www.4art.ee | 44. www.axel.ee | 85. www.dragon.ee |
| 4. www.abtrans.ee | 45. www.axxis.ee | 86. www.eda.ee |
| 5. www.acuba.ee | 46. www.azs.ee | 87. www.fillu.edu.ee |
| 6. www.adeco.ee | 47. www.baas.ee | 88. www.kopli.tln.edu.ee |
| 7. www.adila.ee | 48. www.bakkero.ee | 89. www.palapk.edu.ee |
| 8. www.aff.ee | 49. www.balteco.ee | 90. www.tpl.edu.ee |
| 9. www.aipi.ee | 50. www.keskkula.balhost.ee | 91. www.valgagym.edu.ee |
| 10. www.ajalevi.ee | 51. www.balhotels.ee | 92. www.eestiehitus.ee |
| 11. www.akukon.ee | 52. www.begemotik.ee | 93. www.eeststat.ee |
| 12. www.alarnex.ee | 53. www.bill.ee | 94. www.ehitajad.ee |
| 13. www.alativalmis.ee | 54. www.binekmetall.ee | 95. www.ehitamine.ee |
| 14. www.aleksagor.ee | 55. www.bionext.ee | 96. www.ehitusveod.ee |
| 15. www.alkpipes.ee | 56. www.bitboard.ee | 97. www.ehtas.ee |
| 16. www.almex.ee | 57. www.bim-bom.ee | 98. www.eevali.ee |
| 17. www.altor.ee | 58. www.bonsai.ee | 99. www.elil.ee |
| 18. www.alumiinium.ee | 59. www.boro.ee | 100. www.elsamar.ee |
| 19. www.ambla.ee | 60. www.bruster.ee | 101. www.iseseisev-elu.ee |
| 20. www.anadata.ee | 61. www.bspteam.ee | 102. www.emertek.ee |
| 21. www.andrereklaam.ee | 62. www.buddhabar.ee | 103. www.energo.ee |
| 22. www.apilv.ee | 63. www.camex.ee | 104. www.eni.ee |
| 23. www.apollo.ee | 64. www.cante.ee | 105. www.eradetektiiv.ee |
| 24. www.arens.ee | 65. www.celex.ee | 106. www.eris.ee |
| 25. www.arrak.ee | 66. www.celica.ee | 107. www.essex.ee |
| 26. www.arsenal.ee | 67. www.cerbos.ee | 108. www.estcard.ee |
| 27. www.artig.ee | 68. www.chemie.ee | 109. www.estmer.ee |
| 28. www.kti.artun.ee | 69. www.chr.ee | 110. www.etdm.ee |
| 29. www.arvamus.ee | 70. www.cinderellasalong.ee | 111. www.eurolabo.ee |
| 30. www.arvutikaubad.ee | 71. www.cityguide.ee | 112. www.eurox.ee |
| 31. www.arx.ee | 72. www.compica.ee | 113. www.excelsior.ee |
| 32. www.mac.arx.ee | 73. www.consumetric.ee | 114. www.eye.ee |
| 33. www.asm.ee | 74. www.conteche.ee | 115. www.fifaa.ee |
| 34. www.atelier.ee | 75. www.corelli.ee | 116. www.filosoft.ee |
| 35. www.atvrent.ee | 76. www.dancemix.ee | 117. www.finestmedia.ee |
| 36. www.audentes.ee | 77. www.dataconsult.ee | 118. www.finissage.ee |
| 37. www.audrurfarm.ee | 78. www.denera.ee | 119. www.fixtec.ee |
| 38. www.aulekirjastus.ee | 79. www.denfa.ee | 120. www.fms.ee |
| 39. www.authenticmodels.ee | 80. www.developer.ee | 121. www.folio.ee |
| 40. www.eleven-auto.ee | 81. www.diagnostika.ee | 122. www.folklore.ee |
| 41. www.automaaklerid.ee | 82. www.digitiger.ee | 123. www.fotoarhiiv.ee |

| | | |
|---------------------------------|---------------------------------|--------------------------------|
| 124. www.friedeman.ee | 179. www.krapesk.ee | 234. www.nordstad.ee |
| 125. www.fute.ee | 180. www.krastoehtus.ee | 235. www.openoffice.offline.ee |
| 126. www.gallery.ee | 181. www.kristmet.ee | 236. www.omakodumaja.ee |
| 127. www.nubluka.girls.ee | 182. www.kubija.ee | 237. www.on24.ee |
| 128. www.grill.ee | 183. www.kuuste.ee | 238. www.unginst.org.ee |
| 129. www.haaber.ee | 184. www.kvleht.ee | 239. www.osta24.ee |
| 130. www.habasit.ee | 185. www.ehospehos.kvn.ee | 240. www.pa.ee |
| 131. www.harjuaknad.ee | 186. www.laptop.ee | 241. www.paatsa.ee |
| 132. www.hiedel.ee | 187. www.lcreval.ee | 242. www.paintball.ee |
| 133. www.hitarena.ee | 188. www.leksiko.ee | 243. www.panora.ee |
| 134. www.hmb.ee | 189. www.levst.ee | 244. www.pdc.ee |
| 135. www.hoochimama.ee | 190. www.lexline.ee | 245. www.pergohotels.ee |
| 136. www.hvc.ee | 191. www.liiklusohutuskeskus.ee | 246. www.persona.ee |
| 137. www.ideelabor.ee | 192. www.liin.ee | 247. www.philomed.ee |
| 138. www.illlab.ee | 193. www.linking.ee | 248. www.pimolar.ee |
| 139. www.ilmsalukalakasvatus.ee | 194. www.linnakivi.ee | 249. www.piritaklooster.ee |
| 140. www.iml.ee | 195. www.linnuse.ee | 250. www.pite.ee |
| 141. www.imprest.ee | 196. www.lksa.ee | 251. www.pkelekter.ee |
| 142. www.indigobates.ee | 197. www.llkkatused.ee | 252. www.teiter.planet.ee |
| 143. www.infokas.ee | 198. www.lobe.ee | 253. www.poker.ee |
| 144. www.infowatch.ee | 199. www.locknet.ee | 254. www.pokker.ee |
| 145. www.inspirekeskus.ee | 200. www.loodusmatkad.ee | 255. www.polli.ee |
| 146. www.instant.ee | 201. www.loomakaitse.ee | 256. www.polva.ee |
| 147. www.phon.ioc.ee | 202. www.losttreasure.ee | 257. www.poogen.ee |
| 148. www.isf.ee | 203. www.m39.ee | 258. www.ppi.ee |
| 149. www.jaaboiler.ee | 204. www.madeli.ee | 259. www.ppy.ee |
| 150. www.jargminepeatu.ee | 205. www.maja24.ee | 260. www.2k.pri.ee |
| 151. www.jbauto.ee | 206. www.majas.ee | 261. www.aarne.pri.ee |
| 152. www.jedid.ee | 207. www.margusmaekala.ee | 262. www.aff.pri.ee |
| 153. www.jes.ee | 208. www.markee.ee | 263. www.black.pri.ee |
| 154. www.jets.ee | 209. www.mas.ee | 264. www.cadillac.pri.ee |
| 155. www.jmke.ee | 210. www.matkamees.ee | 265. www.center.pri.ee |
| 156. www.kaasik.ee | 211. www.maxima.ee | 266. www.cool98.pri.ee |
| 157. www.kallaku.ee | 212. www.mazeesti.ee | 267. www.crn.pri.ee |
| 158. www.kalmistud.ee | 213. www.meediatanav.ee | 268. www.curmee.pri.ee |
| 159. www.karitepluss.ee | 214. www.menu.ee | 269. www.daogroup.pri.ee |
| 160. www.kauksipk.ee | 215. www.meos.ee | 270. www.deathaura.pri.ee |
| 161. www.kbyroo.ee | 216. www.merkest.ee | 271. www.demon.pri.ee |
| 162. www.khs.ee | 217. www.metsaomanik.ee | 272. www.dribler.pri.ee |
| 163. www.kiik.ee | 218. www.mibitech.ee | 273. www.edgarprass.pri.ee |
| 164. www.kik.ee | 219. www.miridon.ee | 274. www.erki.pri.ee |
| 165. www.kindrell.ee | 220. www.mit.ee | 275. www.erti.pri.ee |
| 166. www.kingdom.ee | 221. www.mitchel.ee | 276. www.ewteam.pri.ee |
| 167. www.kiwi.ee | 222. www.mmab.ee | 277. www.gabriel.pri.ee |
| 168. www.kmh.ee | 223. www.mospil.ee | 278. www.hakman.pri.ee |
| 169. www.kodutuli.ee | 224. www.mosseliiga.ee | 279. www.henrytiitus.pri.ee |
| 170. www.kolgata.ee | 225. www.mpak.ee | 280. www.hmekkk.pri.ee |
| 171. www.konesko.ee | 226. www.muusikamaa.ee | 281. www.imp.pri.ee |
| 172. www.konfident.ee | 227. www.nagel.ee | 282. www.inge.pri.ee |
| 173. www.konsultant.ee | 228. www.neitsimaarja.ee | 283. www.ingrid.pri.ee |
| 174. www.konteinerveod.ee | 229. www.nek.ee | 284. www.isand.pri.ee |
| 175. www.koplikool.ee | 230. www.netiturg.ee | 285. www.jk.pri.ee |
| 176. www.korpsakala.ee | 231. www.mutionu.nfo.ee | 286. www.kaksik.pri.ee |
| 177. www.korstnapyhkija.ee | 232. www.noltonest.ee | 287. www.kallas.pri.ee |
| 178. www.kosemoobel.ee | 233. www.nonsense.ee | 288. www.karel.pri.ee |

| | | |
|-----------------------------|-------------------------------|-----------------------------|
| 289. www.kordnik.pri.ee | 339. www.pty.ee | 389. www.tevalto.ee |
| 290. www.krx.pri.ee | 340. www.puukaup.ee | 390. www.tevokaup.ee |
| 291. www.lend52.pri.ee | 341. www.raag.ee | 391. www.think.ee |
| 292. www.loodus.pri.ee | 342. www.raamatud.ee | 392. www.thm.ee |
| 293. www.loru.pri.ee | 343. www.rai.ee | 393. www.tkwcy.ee |
| 294. www.love.pri.ee | 344. www.rakvereatp.ee | 394. www.tokanda.ee |
| 295. www.lusikas.pri.ee | 345. www.rannaauto.ee | 395. www.toyotacelica.ee |
| 296. www.maire.pri.ee | 346. www.rapinahaigla.ee | 396. www.travelnet.ee |
| 297. www.manjabes.pri.ee | 347. www.raplakrk.ee | 397. www.triennial.ee |
| 298. www.matukas.pri.ee | 348. www.raudam.ee | 398. www.trt.ee |
| 299. www.max.pri.ee | 349. www.respo.ee | 399. www.trueweb.ee |
| 300. www.muhv.pri.ee | 350. www.revelia.ee | 400. www.trumm.ee |
| 301. www.nucleardawn.pri.ee | 351. www.ringluskeskus.ee | 401. www.trykimaailm.ee |
| 302. www.olar.pri.ee | 352. www.roosasuhkruvatt.ee | 402. www.tsk.ee |
| 303. www.oliver.pri.ee | 353. www.roosikrantsinotar.ee | 403. www.truufoto.tt.ee |
| 304. www.oravaonn.pri.ee | 354. www.rsl.ee | 404. www.chem.ttu.ee |
| 305. www.own.pri.ee | 355. www.rvib.ee | 405. www.humanaaar.ttu.ee |
| 306. www.paavor.pri.ee | 356. www.scandinavia.ee | 406. www.kk.ttu.ee |
| 307. www.pcms.pri.ee | 357. www.seltskond.ee | 407. www.lr.ttu.ee |
| 308. www.pedak.pri.ee | 358. www.semiootika.ee | 408. www.roke.ttu.ee |
| 309. www.peorikkuja.pri.ee | 359. www.sex.ee | 409. www.tuleriit.ee |
| 310. www.pepe.pri.ee | 360. www.sharkline.ee | 410. www.s-tuning.ee |
| 311. www.pilv.pri.ee | 361. www.sillerauto.ee | 411. www.unity.ee |
| 312. www.pokk.pri.ee | 362. www.sirowa.ee | 412. www.uoae.ee |
| 313. www.prof.pri.ee | 363. www.sjs.ee | 413. www.marg.us.ee |
| 314. www.raaliram.pri.ee | 364. www.skinnymob.ee | 414. www.uusvara.ee |
| 315. www.ralff.pri.ee | 365. www.slops.ee | 415. www.v16.ee |
| 316. www.reimers.pri.ee | 366. www.smily.ee | 416. www.vaibla.ee |
| 317. www.riin.pri.ee | 367. www.sojax.ee | 417. www.valgaky.ee |
| 318. www.rl.pri.ee | 368. www.spl.ee | 418. www.valgekass.ee |
| 319. www.rosin.pri.ee | 369. www.srnet.ee | 419. www.vanalinnastudio.ee |
| 320. www.runescape.pri.ee | 370. www.staatus.ee | 420. www.vapotec.ee |
| 321. www.sandra.pri.ee | 371. www.mhes.start.ee | 421. www.veebik.ee |
| 322. www.scifi.pri.ee | 372. www.start.ee | 422. www.verloo.ee |
| 323. www.si.pri.ee | 373. www.stoltsen.ee | 423. www.vermonton.ee |
| 324. www.svenman.pri.ee | 374. www.stunts.ee | 424. www.via.ee |
| 325. www.taimar.pri.ee | 375. www.sunshine.ee | 425. www.crjg.vil.ee |
| 326. www.texmex5.pri.ee | 376. www.sven.ee | 426. www.stat.vil.ee |
| 327. www.themks.pri.ee | 377. www.svsmetall.ee | 427. www.viljandiapteek.ee |
| 328. www.tume.pri.ee | 378. www.taarapunkt.ee | 428. www.viljandibowling.ee |
| 329. www.twist.pri.ee | 379. www.tallinn.ee | 429. www.villa.ee |
| 330. www.unit.pri.ee | 380. www.talvakas.ee | 430. www.vipprojekt.ee |
| 331. www.vecna.pri.ee | 381. www.tammik.ee | 431. www.viskari.ee |
| 332. www.xbs.pri.ee | 382. www.vilde.tammsaare.ee | 432. www.vms.ee |
| 333. www.zaajats.pri.ee | 383. www.tao.ee | 433. www.wikman.ee |
| 334. www.zeus.pri.ee | 384. www.tartublues.ee | 434. www.wildevents.ee |
| 335. www.ziim.pri.ee | 385. www.tat.ee | 435. www.wildmedia.ee |
| 336. www.aik-projekt.ee | 386. www.tehvandi.ee | 436. www.ywam.ee |
| 337. www.ideo-projekt.ee | 387. www.terviklik.ee | 437. www.zepp.ee |
| 338. www.proparts.ee | 388. www.terviseportaal.ee | |