

# HY:n Elektronisen julkaisupankin saamat dokumentit

Timo Hellgren

3. helmikuuta 1999

## 1 Pro gradu -tutkielmat

### 1.1 Eläinlääketiede

#### 1.1.1 Mikko Kananen

Mikko Kanasen pro gradu saatiin tiedekunnan yhdyshenkilöltä valmiina PDF-tiedostoina. Liite- ja kirjallisuussivut tosin näyttivät puuttuvan. Lisäksi PDF-tiedostossa oli jokin 'bookmark error' -teksti. Poistin tämän Adoben Exchangella. Yhdistin lisäksi sisällysluettelotiedoston varsinaiseen päätiedostoon.

Tein palvelimelle HTML-sivun, josta pääsee työhön ja siihen liittyviin tiivistelmiin.

### 1.2 Kasvatustiede

#### 1.2.1 Tuula Kurkinen

Tuula Kurkisen pro gradu saatiin tiedekunnan yhdyshenkilöltä Word 6.0/95 muodossa. Mukana tuli saatetietolomake, jossa väitettiin tiedoston olevan WordPerfect 6.1 -muotoa.

Kokeilin muutamaa ohjelmaa, jotka muuntavat tekstinkäsittelyohjelmien tiedostoja HTML:ksi. Paras tähän asti on ollut HTMLTrans. Muut ohjelmat eivät toimineet. HTMLtrans osasi jakaa Word-tiedoston osiin. Tässä se hyödyntää Wordin tyylejä, joten lopputulos riippuu kirjoittajan taidoista. Jouduin kuitenkin jatkaamaan editointia käsin, sillä kirjoittaja ei ollut hyödyntänyt tyylejä aivan loppuun asti. Hän oli kuitenkin tehnyt taulukot Word-tilukkoina, joten näitä ei tarvinnut alkaa tehdä käsin HTML-versioon.

HTMLtrans ei ole aivan täydellinen, sillä editointia joutuu tekemään vielä käsin. Nyt on hieman vaikea arvioida, sujuiko työ nopeammin kuin jos se olisi käännetty HTML:ksi Wordista käsin, sillä alkuperäinen teksti hyödynsi tyylejä ja taulukoita tavalla, joka nopeuttaa joka taupauksessa konversiota.

PDF-version tekemisessä jouduin editoimaan alkuperäistä dokumenttia hieman. Sivutus ja muotoilut nimittäin riippuvat Wordissa käytössä olevasta tulostimesta, enkä tiedä millä tulostinajurilla Kurkinen oli tulostanut työn itselleen. Lähinnä alussa olevia tiivistelmälomakkeiden kirjasinkokoa piti pienentää, jotta lomake mahtuisi yhdelle A4-arkille käytössäni olevien tulostimien marginaaliastusten mukaisesti.

Pari kaaviota oli tallennettu tiedostoon mukaan. Näiden laatu kärsi PDF-versiossa, koska ne olivat Wordissa ilmeisesti bittikarttamuodossa. Piirsin kuvat uudelleen eräällä ilmaisella vektorigrafiikkaohjelmalla ja tallensin ne EPS-muotoon. Näin kuvat säilyvät tarkkana resoluutiosta ja suurennossuhteesta riippumatta. Toisesta kuvasta tosin tuli hieman isompi kuin alkuperäisestä, joten sivutus muuttui jälleen hieman.

Tein palvelimelle erillisen HTML-sivun, josta pääsee PDF- ja HTML-versioihin.

## **1.3 Maantiede**

### **1.3.1 Erkki Ervast**

Erkki Ervast lähetti pro gradu -tutkielmansa HTML-tiedostoina, jotka oli tehty Microsoft FrontPage 3.0:lla. Työn kuvitus oli GIF-tiedostoina.

Editoin tiedostoja Netscapen Composerilla, sillä linkit eivät ainakaan Netscape Communicatorissa toimineet.

### **1.3.2 Veli-Pekka Ihamäki**

Veli-Pekka Ihamäki lähetti pro gradu -tutkielmansa Adobe Pagemaker 5.0 -tiedostoina, jotka oli tehty Macintoshissa. Kuvituksena olevat kaaviot ja kartat olivat EPS-muodossa. Lisäksi hän lainasi käyttööni paperitulosteen kyseisestä työstä.

ELEKTRA-projektilla on käytössään Power PC-tietokone, johon on asennettu Pagemaker 6.5 (Macintosh-versio). Avasin tiedostot Pagemakerilla, joka automaattisesti päivitti tiedostot ohjelman uuteen versioon. Kuvien linkkaus dokumenttiin piti hoitaa vielä käsin. Suurimmat ongelmat tässä työvaiheessa johtuivat lähinnä siitä, etten ole käyttänyt Macintoshia tai Pagemakeria aikaisemmin.

PDF-tiedostot voidaan tulostaa suoraan Pagemakerista. Tässä ei ollut erityisiä ongelmia. Ainoastaan kuvien osalta piti muuttaa asetuksia niin, että niiden resoluutio vastaisi tulostuslaatua. Tämä tarkoittaa sitä, että PDF-tiedostosta tuli varsin

kookas, yli 12 Mt.

Samoin HTML-tiedostot voidaan tulostaa Pagemakerista käsin. Tällöin teos voidaan jakaa useampaan HTML-tiedostoon. Pagemakerin tyylien muuntuminen tietyksi HTML-tyyliksi voidaan määritellä asetuksista. Normaaliasetuksilla Pagemaker pyrkii tekemään sivusta täsmälleen alkuperäisen näköisen. Poistin kuitenkin tämän asetuksen, sillä alkuperäisen taiton tiukka noudattaminen WWW-ympäristössä ei tunnu tarpeelliselta. Mukana olevan kuvituksen muuntamisen sopivaan formaattiin hoituu Pagemakerilta automaattisesti. Valitsin kuvaformaatiksi GIF:n, sillä teoksen kuvitus oli piirroksia ja kaavioita.

Jatkoin HTML-tiedostojen käsittelyä Netscapen Composerilla. Pagemaker ei lisää automaattisesti sivujen välistä navigointia. Lisäksi joitakin muotoiluja piti tehdä käsin. Dokumentissa käytössä ollut kirjasin muuntui Timesiksi. Palautin kuitenkin alkuperäisen groteski-kirjasimen (tässä tapauksessa Macin Helvetica, jota Windowsissa vastaa Arial), joka varsinkin näytöllä on helppolukuisempi.

PageMaker ei tehnyt kuvituksena olleista kartoista kyllin suuria, jotta niitä olisi näytöllä ollut sanottavaa hyötyä. Niinpä muutin EPS-kuvat GIF-kuviksi itse uudestaan. Muiden ohjelmien puutteessa avasin EPS-kuvan Ghostviewillä ja muunsin sen Windowsin bittikartaksi (BMP). Tämän avasin puolestaan Paint Shop Pro:lla, pienensin kuvaa alkuperäisestä hieman ja muutin GIF-muotoon. Paint Shop Pro ei valitettavasti osaa lukea suoraan Postscript-kuvia, joten kuvat piti muuttaa ensin bittikartoiksi.

### **1.3.3 Timo Arvonen**

Timo arvonen toimitti pro gradu -tutkielmansa Word97-tiedostoina. Kansisivu, sisällysluettelo, liitteet ja jokainen luku olivat omina tiedostoinaan. Sisällysluettelo oli tehty käsin, mutta muuten työ käytti Wordin tyylejä.

HTML-muunnosta ei tarvinnut juurikaan korjata käsin, joten alkuperäiset tiedostot olivat ilmeisesti optimaaliset muunnosta ajatellen. Kuvitusta korjailin kyllä käsin. Word97 muunsi kuvituksen GIF-muotoon, mutta resoluution huonontaminen teki karttojen tekstit lukukelvottomiksi. Kirjoitin karttojen tekstit uudelleen Paint Shop Pro -ohjelmalla.

PDF-tiedostojen teossa ei ollut ongelmia. Tein ensin muunnoksen jokaisesta erillisestä tiedostosta Acrobat Distillerillä ja liitin ne sitten yhteen yhdeksi PDF-tiedostoksi Acrobat Exchangella.

## **1.4 Matematiikka**

### **1.4.1 Ville Hakulinen**

Ville Hakulisen pro gradu -tutkielma saatiin L<sup>A</sup>T<sub>E</sub>X-tiedostona.

Muutin  $\LaTeX$ -tiedoston  $\LaTeX2e$ -tiedostoksi ja käänsin sen dvi-tiedostoksi, josta tein Postscript-tiedoston dvipsillä. Koska käytän mielummin tavallista  $\TeX$ :iä kuin  $ML\text{-}\TeX$ :iä, oli tavutuksen vuoksi otettava käyttöön T1-merkistö eli EC-fontit. Jotta PDF-tiedosto olisi hyvän näköinen olisi käytettävä Postscript-fontteja. Imuroin verkosta Bluesky:n Postscript-versiot Computer Modern fonteista, jotta matemaattiset kaavat voitaisiin ladata myös Postscriptillä. Koska kuitenkin nämä Postscript-fontit noudattavat amerikkalaista OT1-koodausta, otin käyttöön  $\ae$ -makropakkauksen, joka mahdollistaa T1-koodauksen käyttämisen Computer Modern Postscript-fonttien kanssa. Näillä muutoksilla gradusta saatiin kaunis PDF-tiedosto, jossa ei käytetty bittikarttafontteja.

HTML-version tein  $\TeX4ht$ -ohjelmalla. Sillä voi valita, missä muodossa matemaattikka esitetään. Valitsin HTML 3.2:n ja GIF-kuvat. Käännös sentään onnistui, vaikka alkuperäinen  $\LaTeX$ -koodi ei aivan noudata  $\LaTeX$ :n normaaleja käytäntöjä. Jouduin kuitenkin käymään HTML-koodin läpi ja korjaamaan joitakin virheitä käsin. Samalla tein muutaman kuvan lisää ja korjasin joitakin epäonnistuneen tuntuisia kuvia Paint Shop Prolla. Ongelmat johtuivat osin dokumentin hyvin matemaattispuolisesta  $\LaTeX$ -koodista sekä siitä, että  $\TeX4ht$ :n automaattikkaa olisi pitänyt säätää vielä jonkin verran enemmän.

#### 1.4.2 Kari Miettinen

Kari Miettisen pro gradu -tutkielma saatiin Postscript-tiedostona ja tiivistelmämake  $\LaTeX$ -tiedostona.

Käänsin tiivistelmämakeen HTML-tiedostoksi TTH-ohjelmalla ja käsin editoimalla Netscape Composerilla.

Koska tutkielma oli saatu ainoastaan Postscript-tiedostona, oli sen konvertoimisessa odotettavissa vaikeuksia, mikäli ei tyydyttäisi pelkän Postscript-tiedoston julkaisemiseen. Oletettavasti tekstiä lukevan yleisö koostuu enimmäkseen  $\LaTeX$ :n käyttäjistä, joilla on käytössä ohjelmat Postscript-tiedostojen lukemiseen.

PDF:ksi muuttamisessa on ongelmana matemaattiset kaavat. Vaikutti siltä, että teksti käytti Postscript-fontteja, koska nämä näyttivät myös PDF-tiedostossa teräviltä. Kaavoissa taas on käytetty  $\TeX$ :n omia bittikarttafontteja ja nämä tulostuivat epäterävinä näyttöön. (Postscript-tiedostossa kaavat näkyvät terävinä.) Käytin muunnokseen sekä Ghostscriptin PDFWriteriä että Adoben Distilleriä. Laadun kannalta ei ollut huomattavia eroja. Distillerin tekemä PDF-tiedosto oli n. 200 kilotavua pienempi kuin Ghostscriptin. Tämä johtuu siitä, että Ghostscript ei tue LZW-pakkausta tekijänoikeudellisista syistä johtuen.

Toistaiseksi näyttää siis siltä, että itse gradu on julkaistava verkossa Postscript-tiedostona.

### 1.4.3 Jussi I. Pehkonen

Jussi Pehkosen pro gradu -tutkielma saatiin GNUzipillä ja tarilla pakattuina  $\LaTeX$ -tiedostoina. Mukana oli myös  $\text{BIB}\TeX$ -tiedostoja.

Päivitin gradun noudattamaan  $\LaTeX$ 2e-formaattia. Lisäsin lisäksi käskyn käyttää Postscript-fontteja. Käänsin dokumentin  $\LaTeX$ :lla dvi-tiedostoksi, siitä dvipsillä Postscript-tiedostoksi ja Adobe Distillerillä PDF-tiedostoksi. Käännös onnistui muuten hyvin, mutta otsikossa olevat kaksi peräkkäistä f-kirjainta, jotka normaalisti muodostavat ff-ligatuurin, muuttuivat ylöspäin osoittavaksi nuoleksi. Syynä näyttäisi olevan kapiteelikirjasin, jota on käytetty otsikossa. Tekstissä sana ”differentiaali” tulostui aivan oikein ff-ligatuurilla. Ongelma katosi kun purin ligatuurin panemalla toisen f-kirjaimen ryhmytyssulkujen sisään:  $\text{f}\{f\}$ .

Distiller kuitenkin aiheutti toisen ongelman. Gradussa oli yläreunaan merkitty luvun nimi ja numero, mutta distillerillä tehdyssä PDF-tiedostossa tätä ei jostain syystä ollut. Ghostscriptin tekemässä PDF-tiedostossa oli, mutta fontit tulostuivat rumasti bittikarttakuvina.

Päätin kokeilla  $\text{pdf}\LaTeX$ :ia PDF-tiedoston tuottamiseen. Ensin piti säätää hieinan ohjelman asetustiedostoja, jotta Postscript-fontit löytyisivät. Lopputulos oli kuitenkin parempi näin kuin Distillerillä tehtäessä. Sivujen yläreunatekstit näkyivät ja PDF-tiedoston koko oli kylläkin suurempi kuin jopa Postscript-tiedoston, mutta tämä saattoi johtua siitä, etten ollut käyttänyt  $\text{pdf}\LaTeX$ :issa lainkaan pakkaamista. Kun nostin pakkauksen määräksi '8' (tiedostossa `pdf.tex.cfg` arvo `compresslevel`), tuli PDF-tiedoston kooksi enää yli 400 kt (vastaa Postscript-tiedosto on yli 600 kt).

HTML-version tein  $\text{T}\text{E}\text{X}4\text{ht}$ -ohjelmalla. Käytin ohjelman oletusarvoja, jotka muuntavat dokumentin HTML 4.0 standardille. Kaikki ulkoasumäärittelyt tehdään CSS-tyyleillä, joiden niminä on LaTeX-fonttien nimiä. Tämän täytyy johtua siitä, että muunnos tehdään pääosin DVI-tiedostosta eikä `tex`-tiedostosta. Luvun 4 kohdalla jotain meni kuitenkin pieleen. Dokumentti kyllä meni ohjelmasta läpi, mutta loppuosassa oli pahoja käännösvirheitä, jotka korjasin käsin. Lisäksi lisäsin käsin muutaman uuden kuvan.  $\text{T}\text{E}\text{X}4\text{ht}$  ei esim. kääntänyt murtolukuja oikein. Aikaa korjailuihin käytin tehokasta työaikaa lähes pari viikkoa. HTML-version tekeminen matematiikan opinnäytteistä näyttää siis vaativan liikaa käsityötä ja aikaa.

## 1.5 Perinnöllisyystiede

### 1.5.1 Nina Korsisaari

Nina Korsisaaren pro gradu saatiin työn ohjaajalta valmiina PDF-tiedostona. Siirsin työn palvelimelle. HTML-tiedostoa ei kannata tehdä PDF-tiedostosta, joten

työstä on ilmeisesti tyydyttävä pelkkään PDF-versioon.

## **2 Väitöskirjat**

### **2.1 Lääketiede**

#### **2.1.1 Kimmo Wartiovaara**

Kimmo Wartiovaaran väitöskirja saatiin korpulla Word97-tiedostona ja PageMaker 6.5 -tiedostoina CD-R-levyllä.

Käänsin Word97-tiedoston HTML-tiedostoksi Word 97:lla ja jatkoin dokumentin käsittelyä Netscape Communicatorin editorilla. Jaoin dokumentin osiin lukujen mukaan ja lisäsin navigointipalkit ylä- ja alapuolelle sekä dokumentin sisäiset linkit. Tekstin ulkoasuun ei juurikaan tarvinnut puuttua. Ainoastaan luettelo lyhennyksistä kaipasi editointia. Luettelo oli tehty kaksipalstaiseksi tabulaattorilla. Tämä piti luonnollisesti muuttaa HTML:ssä taulukoksi.

### **2.2 Maantiede**

#### **2.2.1 Pirjo Hellemaa**

Pirjo Hellemaan väitöskirja saatiin korpulla WordPerfect 5.1-, Excel-, ja ASCII-tiedostoina.

Avasin WordPerfect-tiedostot WordPerfectin versiossa 8 ja tallensin niistä HTML-versiot. Samoin avasin Excel-taulukon Excel 97:ssa ja tallensin sen HTML-tiedostoksi. Seuraavaksi aloin tarkastella ja editoida HTML-tiedostoja Netscape Communicatorin Composerissa.