

TTH:n käyttö HTML:n tuottamiseksi

Timo Hellgren

11. syyskuuta 1998

1 Johdanto

TTH on ilmainen ohjelma, jolla \LaTeX - ja \TeX -dokumentteja voidaan muuntaa HTML:ksi. Kokeiluissa käytettiin ohjelman versiota 1.56.

2 Kokeiluaineisto

Kokeiltavana olivat seuraavat dokumentit:

1. Kari Miettisen pro gradu -tutkielma *Numerical methods for unconstrained minimization: an integrated computational environment*. Tästä dokumentista sain ainoastaan tiivistelmälomakkeen `.tex`-tiedoston sekä itse tekstin Postscript-tiedostona.
2. Ville Hakulisen pro gradu -tutkielma *Kvantifointi määriteltävien relaatioiden yli*. Tästä dokumentista on käytössä kääntämätön \LaTeX -tiedosto (`.tex`) sekä Postscript-tiedosto.
3. Jussi I. Pehkosen pro gradu -tutkielma *Dirichlet'n reuna- ja ominaisarvo-tehtävät eräälle ei-hypoelliptiselle differentiaalioperaattorille*. Dokumentista on käytössä \LaTeX - ja \BIBTeX -tiedostot.

3 Case Miettinen

Kokeilin kääntää Kari Miettisen pro gradun tiivistelmälomakkeen HTML:ksi. Annoin Windows 95:n MS-DOS shellissä käskyn:

```
tth <abstract.tex >abstract.html
```

Ohjelma kyllä teki jonkinlaisen HTML-tiedoston, mutta sisältö ei ollut aivan oikea. TTH osasi tehdä kyllä tiivistelmälomakkeesta taulukon, mutta osa tekstistä jäi puuttumaan, esim. itse tiivistelmäteksti puuttui kokonaan.

Tiivistelmälomakkeen taulukko oli tehty \LaTeX :n `tabular`-ympäristöllä, jota manuaalin mukaan TTH:n kyllä pitäisi ymmärtää. Lomake muuten on tehty \LaTeX :n vanhalla versiolla, ei $\LaTeX2e$:llä, mutta tällä ei ole merkitystä muunnoksen kannalta.

Seuraavaksi aloin tutkia TTH:n virheilmoituksia. Tein lisäksi kokeiluversion tiedostosta nimelle `testi.tex`. TTH ei ymmärtänyt alussa olevia mm. sivun koon määrittäjiä, jolla ei HTML-dokumentissa ole merkitystä. Kommentoin nämä pois. Lisäksi poistin groteskin kirjasimen määrittäjät, joita TTH ei ymmärtänyt, vaikkakin osan tekstiä se osasi silti kääntää tästä huolimatta.

Tein käynnöksen uudelleen. Ulkoasussa ei ollut muutoksia, mutta virheilmoitusten luettelo oli supistunut. TTH ei ymmärrä näköjään myöskään `\parbox`-käskyä. **Bingo!** `\parbox`:n parametreina olivat juuri ne tekstit jotka eivät tulostuneet HTML-tiedostoon. Poistin `\parbox`-käskyt. Näiden poistamisessa pitää kuitenkin olla hyvin huolellinen, jottei sekoitaisi koko taulukkoa ja tiedostoa. Jouduin tekemään työn muutamaan kertaan uudestaan. Muunnos ei kuitenkaan onnistunut. Kun poistin jälkimmäisen `\parbox`:n itse tiivistelmätekstistä, HTML-tiedosto sekosi täysin.

Nyt oli siis ongelmana edelleen tiivistelmätekstin saaminen näkyviin. Kopioin tiivistelmätekstin erilliseen tiedostoon `tiivis.tex`, ja käänsin sen TTH:lla omaksi HTML-tiedostokseen. Samalla muutin `eqnarray*`-ympäristön tavalliseksi \TeX :n matemaattiseksi kaavaympäristöksi (`$$`-käskyillä). Teksti tuli muuten oikein, mutta matemaattiset kaavat eivät tulostuneet kursiivilla ja mustia palloja käyttävä lista ei tulostunut oikein: listasymboliksi tuli pallon sijasta kunkin rivin ensimmäinen kirjain. Kursiivi saadaan matemaattisiin kaavoihin TTH:n parametrilla `-i`. Ohjelma ajetaan siis seuraavalla käskyllä:

```
tth -i <tiivis.tex >tiivis.html
```

Listan vikana oli se, että sen aloittava käsky `\begin{itemize}` olikin kirjoitettu muotoon: `\begin {itemize}`. TTH ei ymmärrä välilyöntiä `\begin`-käskyn ja sen parametrin välissä. Näillä korjauksilla tiivistelmäteksti tulostui oikein tiedostoksi `tiivis.html`.

Nyt saatoin yhdistää nämä kaksi tiedostoa HTML-editorissa. Samalla taulukkoa oli syytä hieman korjailta paremman näköiseksi. HTML-editorina käytin Netscape Communicatorin editoria.

4 Case Hakulinen

Ville Hakulisen gradun lähdekoodi noudatti vanhaa \LaTeX :ia, ei nykyistä $\LaTeX 2\epsilon$:tä. Kokeilin gradun kääntämistä HTML:ksi käskyllä:

```
tth -i <ville_hakulinen.tex >gradu.html
```

Kuten arvata saattaa, käänös ei onnistunut. Näin aluksi suurimmat ongelmat näyttivät olevan `tabbing`-ympäristössä, jota TTH ei tunnista. Lisäksi TTH ei näyttä pystyvän tulostamaan kaikkia tarvittavia symboleja. Mahdollisesti syystä, että `symbol`-kirjasimessa ei ole kaikkia tässä dokumentissa tarvittavia symboleja.

Päätin kokeilla dokumentin kääntämistä HTML:ksi `LaTeX2HTML`:llä.

5 Case Pehkonen

Jussi I Pehkosen gradun lähdekoodi noudattaa vanhaa \LaTeX :ia, ei nykyistä $\LaTeX 2\epsilon$:tä. Koska minulla ei ollut Postscript-tiedostoa gradusta, oli teksti ensin käännettävä \LaTeX :illa. Gradu kyllä kääntyi ilmeisesti aivan oikein, paitsi ä- ja ö-kirjainten osalta, jotka jäivät pois. Syynä oli se, että dokumentti käytti Computer Modern -kirjasimia, mutta ääkkösiä ei oltu kirjoitettu muodossa `\"a` tai `\"o`, joten ainaakaan minun \LaTeX -formaattissani ääkköset eivät tulostu oikein.

Itse käytän $\LaTeX 2\epsilon$:tä, jossa suomea voidaan kirjoittaa kyllä suoraan Windows-koodauksella, mutta kirjasinkoodauksena käytetään \LaTeX :ssa T1-koodausta. Tätä koodausta ei voi kuitenkaan käyttää vanhassa \LaTeX :ssa, niinpä muutin dokumentin alun vastaamaan uutta standardia. Tekstissä on kuitenkin jossain vaiheessa käytössä symboli `\$box\$`, jota $\LaTeX 2\epsilon$ ei tunnistanut. Kyseinen symboli kuuluu ns. \LaTeX -symboleihin, jotka eivät automaattisesti enää sisälly nykyiseen \LaTeX :iin, vaan `latexsym`-makropakettiin, joka pitää ladata dokumentin alussa.

Lisäsin siis dokumentin alkuun käskyt:

```
\usepackage[finnish]{babel}
\usepackage[T1]{fontenc}
\usepackage[ansinew]{inputenc}
\usepackage{latexsym}
```

ja muutin käskyn:

```
\documentstyle[11pt,finnish]{article}
```

$\LaTeX 2\epsilon$ muotoon:

```
\documentclass[11pt]{article}
```

Näiden muutosten jälkeen dokumentti kääntyi mukisematta (muutamaa tavutusongelmista johtuvaa ylipitkää riviä lukuunottamatta) .dvi-tiedostoksi ja siitä dvips:llä Postscript-tiedostoksi.

Seuraavaksi oli aika kokeilla TTH:ta. Annoin MS-DOS shellissä käskyn:

```
tth -i <gradu.tex >gradu.html
```

Ei aivan onnistunut. TTH tukehtui jossain vaiheessa omaan mahdottomuuteensa ja dokumentti jäi kesken. Ääkköset sentään tulivat paikalleen, mutta kaikkia matemaattisia symboleja ei ohjelma osannut kääntää. Lisäksi $\text{BIB}_{\text{T}}\text{E}_{\text{X}}$:n kirjallisuusviittaukset eivät onnistuneet ja paljon muutakin jäi kääntämättä. Päätin jatkaa kokeiluja LaTeX2HTML-ohjelmalla. TTH ei näytä oikein selviävän matematiikan graduista.