



M Ű E G Y E T E M 1 7 8 2
Budapesti Műszaki és Gazdaságtudományi Egyetem
Méréstechnika és Információs Rendszerek Tanszék

Tudományos művek készítése Markdown nyelven

Stein Dániel I. évf., MSc mérnökinformatikus szakos hallgató

Konzulens: Szárnyas Gábor – MIT, dr. Ráth István – MIT

Kritikus rendszerek főspecializáció

Önálló laboratórium 2. összefoglaló

2015/16. I. félév

Tudományos dokumentumok publikációs folyamatának fontos része a dokumentum megfelelő szerkesztése és formázása. A tárgy során feladatom a tudományos dokumentumok publikálásához szükséges munkafolyamat egyszerűsítése volt.

Még a kari TDK és szakdolgozat munkák közt is számos olyan található, mely az ún. WYSIWYG (what you see is what you get) eszközök – többek között Word vagy Pages – használatával készültek. Közre játszhat ebben az is, hogy más megoldások – például a kifejezetten tudományos publikációk készítésére szánt LaTeX – telepítése és használata bonyolult, körülményes és az eredményes használatukhoz előzetes ismeretek ajánlottak.

Az utóbbi években észrevehető a markup, azaz jelölőnyelvek használatának elterjedése egyszerűbb tartalmak esetén. Az egyik leggyakrabban alkalmazott ilyen nyelv a Markdown, mellyel könnyen olvasható és írható dokumentumok készíthetők. A Markdown formátumú dokumentumok átalakítására, fordítására több eszköz is rendelkezésre áll.

A Pandoc – Haskell nyelven írt, nyílt forráskódú és széles körben használt – dokumentumátalakító keretrendszer rengeteg ki- és bemeneti formátum között képes köztes AST (absztrakt szintaxisfa¹) használatával átalakítani. Az egyik bemeneti formátuma a saját nyelvi elemekkel bővített Pandoc Markdown, míg a kimenetei között megtalálható a LaTeX, mely PDF-re fordítható, emellett HTML weboldalak, EPUB fájlok és Word dokumentumok is generálhatók segítségével.

Különböző AST-t átalakító filterekkel elérhető, hogy a tudományos dokumentumokra jellemző tartalmak – idézetek, hivatkozások, táblázatok, ábrák, grafikonok – egyszerűen beilleszthetők legyenek a dokumentumokba a nyelv megkötései mellett. A félév során átalakítottam a tanszéki Markdown formátumú szakdolgozat mintát a filterek (pandoc-crossref, pandoc-placetable...) által nyújtott elemeket használva.

Emellett készítettem egy filtert², mely lehetővé teszi dokumentumrészek hierarchikus hivatkozását, mely nagyobb terjedelmű dokumentumok esetén gyakran használt megoldás az átláthatóság és könnyű kezelhetőség érdekében. (Például bevett szokás a szakdolgozat egyes fejezeit külön fájlokban tárolni és egy fő dokumentumból ezeket hivatkozni.) A filtert a Haskell csomagkezelőjébe is publikáltam³, mely így bárki számára elérhető.

Mivel a LaTeX és a Pandoc telepítéséhez szükséges csomagok (különösen a számos Haskell, Cabal és a pandoc csomag) telepítése nem csupán tárhely-, de időigényes folyamat is, létrehoztam egy ezeket tartalmazó Docker leíró fájlt. A Docker konténer fájlok fordítási tulajdonságát kihasználva – miszerint csupán egyszer szükséges elvégezni a csomagok letöltését és telepítését – nem elhanyagolható mértékű időt spórol a felhasználók számára a megoldás használata. A több gigabájtos telepítő binárisok letöltése és telepítése helyett már előre telepített (a fejlesztő által fordított), kész rendszert tölt le a felhasználó. Ez letöltés után szinte egyből használható, emellett a felhasználó számítógépéről épp oly egyszerűen, maradéktalanul eltávolítható, mint ahogy telepíthető.

¹ <https://hackage.haskell.org/package/pandoc-types-1.12.4.7/docs/Text-Pandoc-Definition.html>

² <https://github.com/steindani/pandoc-include/>

³ <https://hackage.haskell.org/package/pandoc-include>