

# Diplomadolgozat tématerv

Elemző szoftver fejlesztés PDF fájlok vizsgálatához

Balogh Gergő

BAGOADT.SZE

Programtervező Informatikus (Msc)

Témavezető: Dr. Alexin Zoltán

Kelt: 2011. június 29.

## 1. A téma rövid ismertetése

Napjainkban a számítógép általános és egyre több területen történő alkalmazása megköveteli a digitális és nyomtatott dokumentumok egységes kezelését. Az internet elterjedése miatt ezek gyors továbbítása is nélkülözhetetlen. Erre kínál egy megoldást az Adobe Systems Incorporated PDF (*Portable Document Format, Hordozható Dokumentum Formátum*) fájlformátuma.

A PDF képes két dimenziós dokumentumok, vektorgrafikák, (esetleg három dimenziós beágyazott objektumok) eszközfüggetlen tárolására, és továbbítására a hálózaton keresztül. Ez a formátum egyre szélesebb körben használt.

Főként e tények alapján esett a választásom, ezen szakdolgozat témára. Bár a formátum főként a dokumentumok biztonságos tárolására szolgál, bizonyos esetekben szükséges lehet a benne tárolt adatokhoz hozzáférni. Például keresés során; gondolok itt többek közt az internetes keresőmotorokra is, melyek számára nyilván könnyebbség ha a fájlt már előzőleg feldolgozták.

Továbbá tapasztalataim szerint igen kevés embernek van lehetősége, a már kész PDF dokumentumok szerkesztésére. Mégis sokszor szükséges, hogy az ezekben található adatok jól kezelhető formában rendelkezésünkre álljanak. Ezért BSC szakdolgozatomban egy visszafejtő program megtervezését és megvalósítását végeztem el.

A témát folytatva, szeretnék egy absztrakciós szinttel feljebb lépni és támogatást nyújtani a további PDF alapú szoftverek kifejlesztésében. Mindezt egy elemző eszközzel biztosítanám.

## 2. A program rövid ismertetése

A programot a **Microsoft Visual Studio 2010 C#** fejlesztő környezetben *C# nyelven* szeretném megvalósítani. A fejlesztés során egyéb ingyenes segédeszközt is használnék (pl.: FreeMind, egyéb UML szerkesztő). A programom fő kimeneti formátuma a **TEX (L<sup>A</sup>T<sub>E</sub>X** fájl), ill. az **XHTML** lenne, így a teszteléshez használnám a szintén ingyenes **MikT<sub>E</sub>X L<sup>A</sup>T<sub>E</sub>X** disztribúciót ill. a **T<sub>E</sub>XMakerX** szövegszerkesztőt.

A megvalósítás során felhasználnám és továbbfejleszteném az általam készített *ReTeX* visszafejtő könyvtárat. Majd erre építenék egy grafikus alkalmazást, mely lehetővé teszi a PDF fájlok elemzését.

A program a syntax highlighting technikát alkalmazva megjelenítené a fájlok belső struktúrá-

ját. A megjelenítés főleg az emberi felhasználást tartja majd szem előtt, így elvégzi az egyes bináris műveleteket (tömörítés, dekódolás) és az eredményüket szöveges, vagy valamilyen vizuális megjelenítés segítségével szemlélteti. Az egyes elemek részletes kivizsgálására is lehetőséget biztosítanak, ez külön (az adott elemre specifikus) nézetben történne.

Szakedolgozatomban részletezném a megvalósítás során szerzett szakmai tapasztalatokat. Külön kitérnék a szabvány általam észlelt hiányosságaira és a visszafejtés ebből fakadóan az elemzés nehézségeire.

A programban szeretném felhasználni a .NET 4.0 *keretrendszer* nyújtotta újításokat a grafikus felület kialakítása során. Ezért a felhasználói felületet nagy részét az XAML jelölő nyelven valósítanám meg, beépítve a különböző transzformációkat és átmeneteket a felhasználói élmény fokozása érdekében.