A programozás alapjai 2.

Véglegesített házi feladat specifikáció

**File titkosító program I.**

Készítette: Albrecht Ádám, O9E6U1

adam.albrecht@edu.bme.hu

Készítés féléve: 2024/25/2

Tartalomjegyzék

[Dokumentációval kapcsolatos teendők 3](#_Toc478243943)

[Informális házi feladat leírás 4](#_Toc478243944)

[Formális házi feladat leírás 4](#_Toc478243945)

[Az elkészítendő szoftver célja 4](#_Toc478243946)

[Elvárások 4](#_Toc478243947)

[Elvárt bemenet 4](#_Toc478243948)

[Elvárt kimenet 4](#_Toc478243949)

[Implementálandó funkciók 4](#_Toc478243950)

[Felhasználói interakció 4](#_Toc478243951)

[Fejlesztéshez használt technológiák 4](#_Toc478243952)

[Egyéb követelmények 4](#_Toc478243953)

[Futtató környezet követelményei 4](#_Toc478243954)

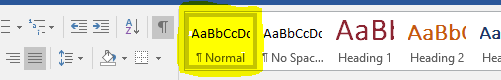
[Célközönség 5](#_Toc478243955)

[Hivatkozások 5](#_Toc478243956)

# Dokumentációval kapcsolatos teendők

Először is, ez a dokumentáció csak egy irányelv, amit örülnék, ha mindenki be tudna tartani annyira, amennyire csak tud. Amennyiben a feladat úgy kívánja, nyugodtan adj hozzá új headeröket.

A piros színnel írt példát elolvasás után **töröld ki** és „Normal” szövegstílussal fogalmazd meg az adott részt a feladatodhoz kapcsolódóan.



1. ábra „Normal” szövegstílus

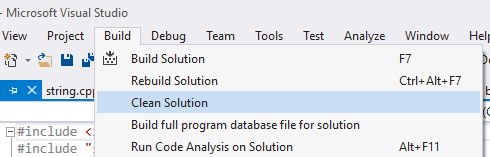
Ezen kívül **ezt az oldalt töröld ki**, mivel nem szerves része a feladatspecifikációnak.

Továbbá a fájl nevében a „NEPTUN” részt, valamint az első oldalon a minta készítőt (Gipsz Jakab, ABC123, gipsz.jakab@minta.hu) **írjd át a saját adataidnak megfelelően**, ezen kívül szintén az első oldal „Feladat cím” részt is **cseréld ki a feladatod címére.**

Ha a fentiekkel elkészültél, **generáld újra a tartalomjegyzéket**, **exportáld ki a dokumentumot PDF formátumban**.

**Ne felejtsd el** az utolsó labor napjának végéig feltölteni zippelve a házi feladatod, ami mindenképp tartalmazza a következőket:

* Az elkészült program (a generálható fájlokat ne tartalmazza: Visual Studio-ban Build menü -> Clean Solution)
* A korábban elküldött házi feladat specifikáció PDF formátumban
* Ez a dokumentáció PDF formátumban



2. ábra Clean Solution

# Informális házi feladat leírás

Mivel még nincs megfelelő, hatékony megoldás morfológiai egyértelműsítőre magyar nyelvhez deep learning technológiával, írok egyet.

# Formális házi feladat leírás

## Az elkészítendő szoftver célja

Ismeretes, hogy nem heurisztikus morfológiai egyértelműsítő még nem létezik az olyan agglutináló nyelvekre, mint például a török, a német vagy éppen a magyar. A szoftverem célja a már meglévő megoldások alapján deep learning technológiákat felhasználva implementálni egy morfológiai elemzés egyértelműsítőt a magyar nyelvre.

## Elvárások

### Elvárt bemenet

A program futtatási argumentumként meg fogja kapni egy morfológiai szóelemző kimenetét, amelyben a korpusz tokenizált szavai külön-külön sorban szerepelnek az egyes lehetséges elemzésekkel tabulátor karakterrel szeparáltan. A neurális hálózat feltanításához szükséges még a helyes elemzés, ez minden esetben a legelső lesz.

### Elvárt kimenet

Egy feltanított hálózat, amely képes lesz megmondani, hogy hány százalék a valószínűsége annak, hogy feltételezve, hogy szóelemzés ablakban az első N-1 jó elemzés, az N. is helyes. A súlyokat fájlba fogom kiírni.

## Implementálandó funkciók

* Szöveges bemenetet feldolgozó algoritmus
* Neurális hálózat architektúra
* Tanítást végző algoritmus
* Tanítást kiértékelő algoritmus
* Neurális hálózat súlyainak fájlba írása

## Felhasználói interakció

A program a futáshoz szükséges hiperparamétereket argumentumként fogja megkapni és a futás során keletkező állapotleíró üzeneteket a standard kimenetre fogja kiírni.

## Fejlesztéshez használt technológiák

A programomat C++11 nyelven írom meg, mivel a kevés bináris overhead miatt várhatóan gyorsan fog lefutni a kódom. Integrált fejlesztőkörnyezetként Microsoft Visual Studio 2017-et fogok használni, emellett az új GitHub integráció segítségével GPLv2 licensszel publikálok.

## Egyéb követelmények

### Futtató környezet követelményei

A programomat Windows környezetre fogom lefuttatni Visual Studio 2017-ben, így a futtatható binárist azon is fogom tudni lefuttatni. Azonban igyekszem csak cross-platform nyelvi elemeket felhasználni a kód írása során, így Linux platformra való fordítás után akár azon is futtatható lesz.

### Célközönség

A programom a magyar természetes nyelvfelismeréssel (NLP – Natural Language Processing) foglalkozó kutatók számára készül.

### Hivatkozások

Google, StackOverflow, stb. linkek.