

A mai magyar helynévi adatbázis feltöltésének munkálatai A projektirányító feladatai

1. ütem

- Ebben az ütemben kerül sor a forrásként szolgáló földrajzinév-tárak anyagának adatbázisban való rögzítésére.
- Ennek során nem történhet adatvesztés, de nem adunk meg a forráshoz képest többletinformációkat sem.
- A munka **településenkénti egységekben** folyik, alapjául az adott település anyagának szövegfájlja szolgál.

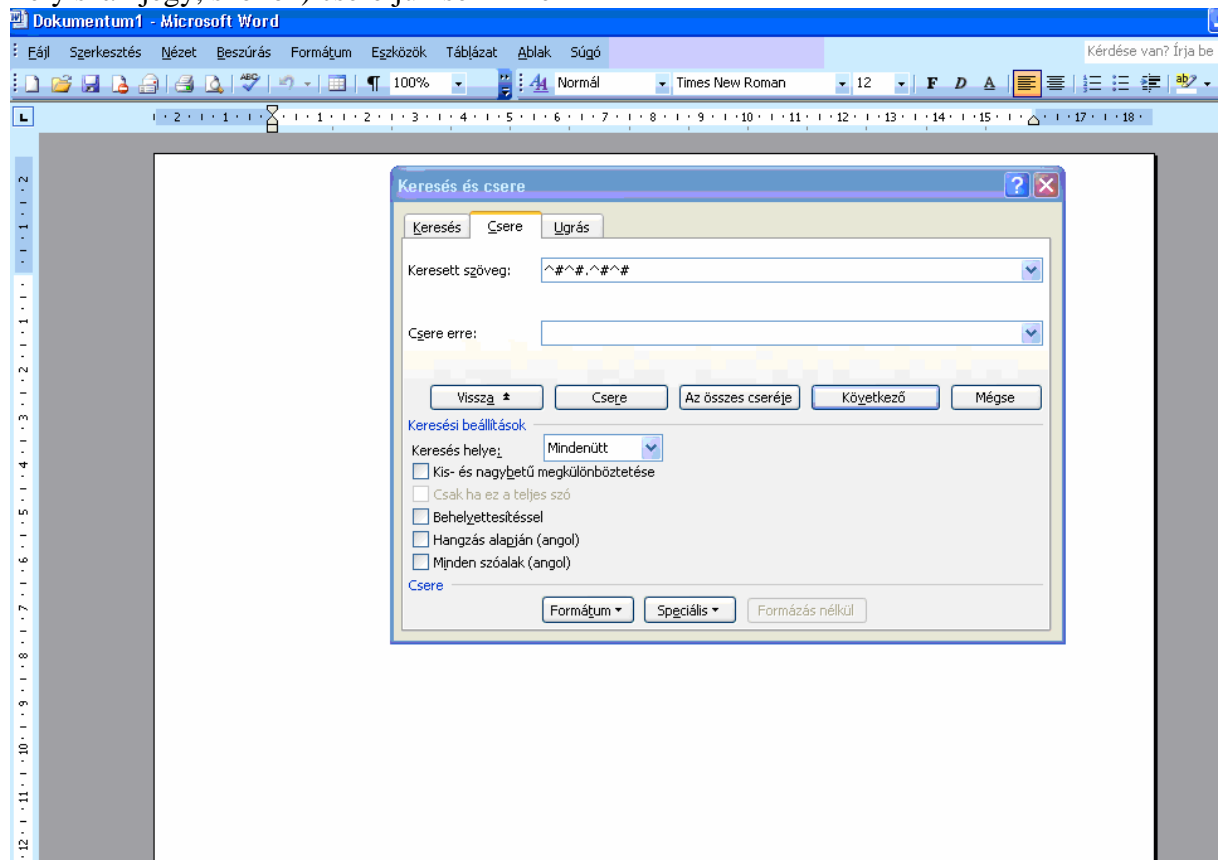
I. Előkészítés

0. Az egyes megyék anyagát az eredeti formában elmentjük (baranyanyers stb. néven), és CD-re is kiírjuk

1. Az alapszövegen végzendő (korrigálási) feladatok (ezt a munkafázist a sorhű szövegfájlban, a teljes megye anyagában célszerű elvégezni)

karaktercserék a fájlban: a nem köznyelvi karaktereket az adatbázisban használt egységes fontkészlet megfelelő karaktereivel helyettesítjük a mellékelt útmutató szerint haladva a szövegfájlban minden sor elején található 4 karakternyi szám és egy szóköz, ennek a funkciója a sorok jelölése volt (ez biztosította tehát a sorhűséget), ám erre a továbbiakban nincs szükség. A megszüntetése:

A négy karakternyi számot (bármely számjegy, bármely számjegy, bármely számjegy, bármely számjegy, szóköz) cseréljük semmire

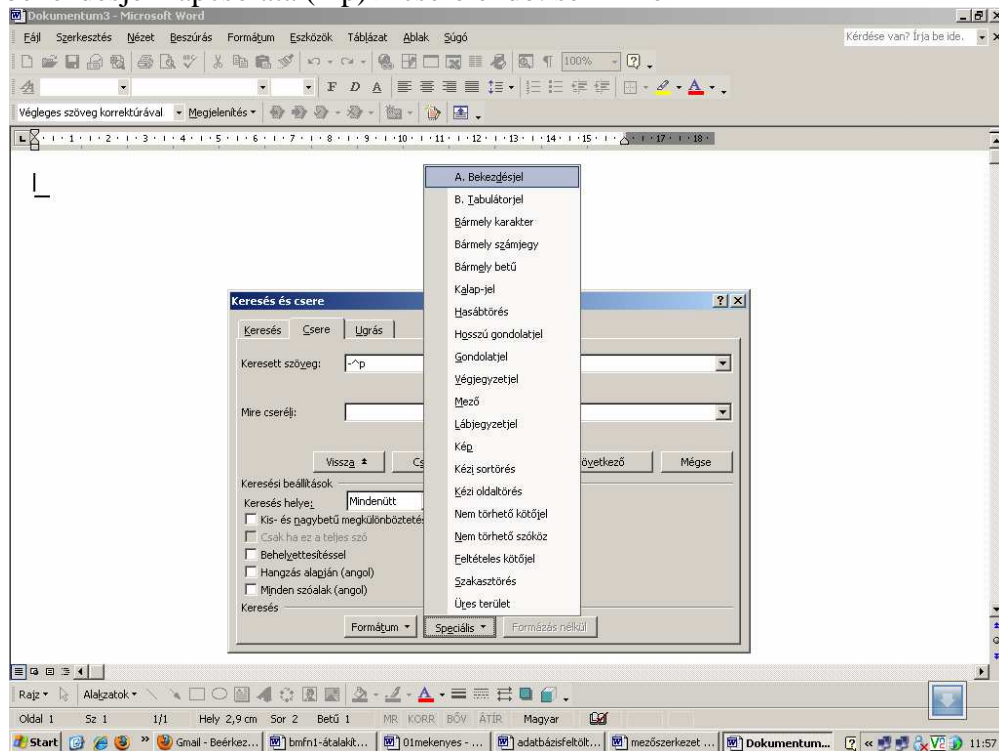


az így átalakított szöveget lementjük egy külön file-ban (baranyajavított stb. cím alatt)

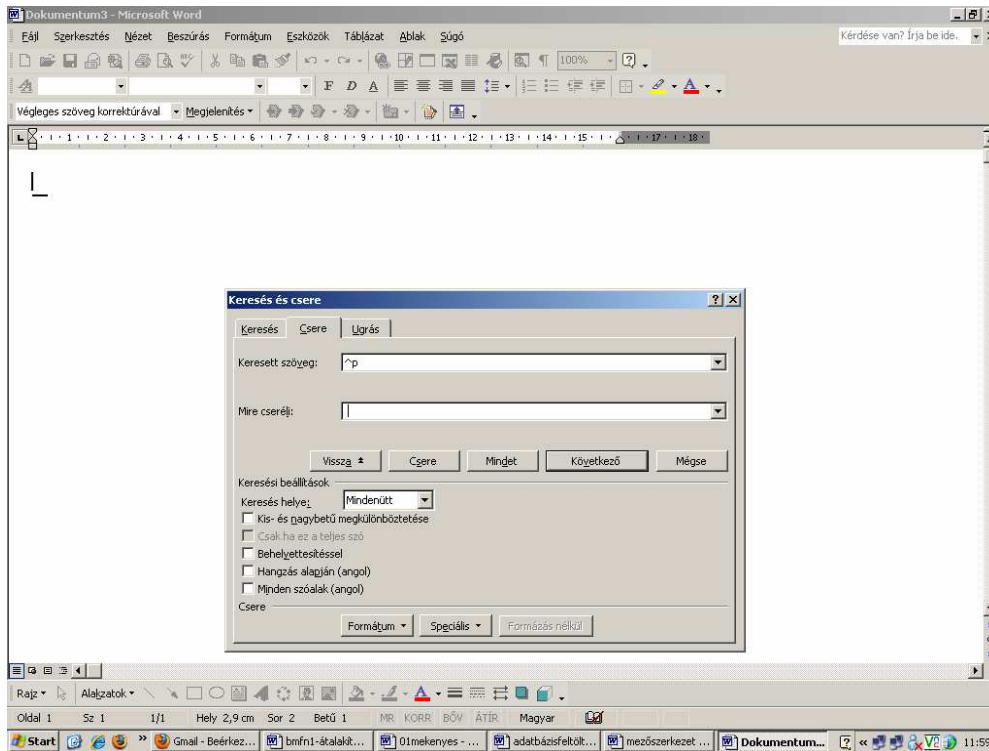
majd ezt települések szerint szétszedjük: minden település egy-egy file-ba kerül még ebben a sorhú változatában (pl. 01mekenyes-sorhú). A sorhúség megtartása azért lényeges, mert a szöveg visszaellenőrzését ebben fogják elvégezni a munkát végzők az adatbázis-készítést megelőzően. Ezeket a fájlokat megyéknént külön könyvtárban (településenkénti/sorhú /baranya /fájlnév) mentjük el, és írjuk CD-re.

2. A szöveg átalakítása folyamatos szöveggé, a sorhúség megszüntetése cserékkel megszüntetjük a sortöréseket:

- és bekezdésjel kapcsolata (-^p) > cserélendő: semmire

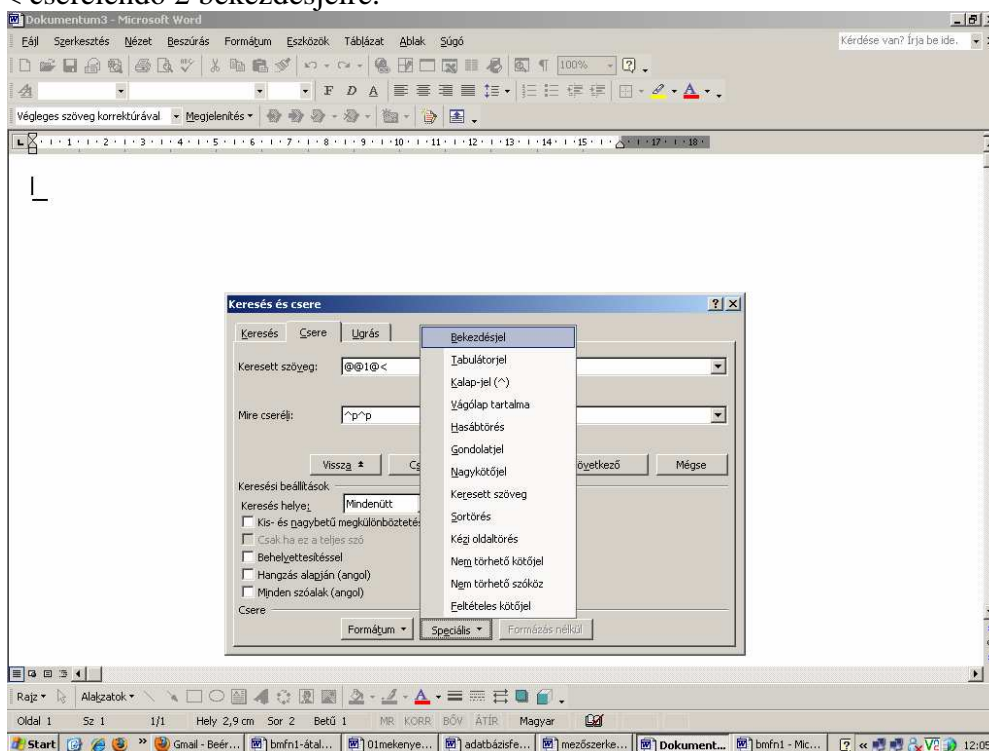


bekezdésjel (^p) > cseré: szóközre



Ezekkel a cserékkel egy folyamatos szöveget kapunk. Ezen belül aztán már csak arra van szükségünk, hogy a szöveg laponként tagolódjon:

A szöveg lapszámok szerinti tagolását a következőképpen végezhetjük el: az eredeti szövegfile-ban a lapszámok ezzel a szerkezettel vannak megadva (Baranya esetében): <P> S95> vagy <-P> S95> (ez pl. a 95. lapot jelöli), vagy (Somogy és Tolna esetében) @@1@<S48>. A lapszámok szerinti tagolás ezért cserével szintén könnyen elvégezhető: </P> S95>, ill. <-P> vagy @@1@< cserélendő 2 bekezdésjelre.



3. **Az adatbázis szempontjából fölösleges információk törlése a szöveg elejéről és végéről:** a település adatai 0. sorszámot kapnak az ezeket (és a róla szóló információkat) bemutató első (most már 0.) bekezdés végétől, a „– T” jeltől kezdve (ahol a terület- és lakosságinformációk vannak) kivágjuk az információkat egészen az **1.** sorszámmal jelölt objektumig, és a települések végén lévő, a forrásokra és az adatközlőkre vonatkozó információkkal együtt külön fájlban mentjük el (pl. 01mekeny-es-informacio), megfelelő könyvtárba (információk/baranya/fájlnév) helyezve és CD-re írva.
4. **A szöveg elejére másoljuk a táblázat 10 oszlopból álló változatát.**
5. a feldolgozandó kötet **anyagközlésre vonatkozó útmutatóját** beszkenelve pdf-formátumban föl kell rakatni a tanszék honlapjára, hogy ezt az adott kötetten dolgozó hallgatók el tudják olvasni a munka megkezdése előtt (mnytud.arts.unideb.hu/kutatás/nyelv- és névtörténeti műhely alá)

Ez az a fázis, ahol az adatbázisok feltöltésével foglalkozók megkapják a file-okat:

1. az adott település sorhú file-ját (pl. 01mekeny-es-sorhú), ez a visszaellenőrzéshez kell (a talált hibákat mindkét file-ban javítják majd az adatbázisokat feltöltők)
 2. ehhez megkapják ennek a nyomtatott változatát is
 3. emellett megkapják az adott település anyagának fénymásolatát is, valamint a feloldandó rövidítésjegyzék másolatát is
 4. valamint megkapják file-ban azt a szövegfile-t, amiben a sorhúséget megszüntettük, és amelynek elejére bemásoltuk a táblázatszerkezetet (pl. 01mekeny-es-ömlesztett). Ebből készülnek ugyanis majd az adatbázisok
6. **A településenként visszaérkező szöveges fájlok (sorhú, ömlesztett) javításainak ellenőrzése, a javítatlan változatok cseréje a megfelelő könyvtárban a javítottra, átírás CD-re.**
 7. **A településenként visszaérkező fájlok közül a táblázatfájl elmentése külön könyvtárban (táblázat/baranya/fájlnév):** (pl. 01mekeny-es-táblázat), CD-re írása.
 8. **Excel adatbázis-kezelőbe történő konvertálás** (egyszerű bemásolással: a Word táblázat oszlopaikat kijelölve Másolás funkció, az Excel munkafüzet 1. cellájába állva Beillesztés funkció). Mentés adatbázis/baranya/fájlnév címen.

Tehát az alábbi könyvtárak létrehozása szükséges:

1. nyersanyagok/baranyanyers
2. javítottanyagok/baranyajavított (jó lesz ez, mert ezek az egész baranyai anyagot tartalmazó file-ok, a hallgatóktól pedig már csak egy-egy település jön vissza)
3. településenkénti/sorhú/baranya/01mekeny-es-sorhú
4. településenkénti/ömlesztett/baranya/01mekeny-es-ömlesztett
5. információk/baranya/01mekeny-es-információ
6. táblázat/baranya/01mekeny-es-táblázat