**Digitálisan született anyagok kezelése**

Műhely az ELTE oktatói részére az ELTE Digitális Bölcsészet Központjában

5 x 45 perc

Oktató: Horváth Péter

**Kurzusleírás:**

A kurzus során a résztvevők megismerhetik a fontosabb szövegformátumokat, a szövegekben való hatékony keresés módszereit, valamint a szöveges adatbázisok egyik fő formátumának, az XML-formátumnak a jellemzőit. A kurzus betekintést ad az automatikus nyelvi elemzés alapvető eljárásaiba, és bemutatja a szövegeket tartalmazó adatbázisok (korpuszok) létrehozásának a főbb lépéseit.

**Tematika:**

* PDF-ek jellemzői, PDF-ek szövegfelismertetése
* Fontosabb szövegformátumok jellemzői
* Szövegekben való keresés reguláris kifejezésekkel
* Az XML-formátum alapjai, TEI XML
* Szövegek automatikus nyelvi elemzése
* Szövegeket tartalmazó adatbázisok (korpuszok) létrehozása