

A SZÖVEGVIZUALIZÁCIÓ ESZKÖZTÁRÁNAK ALKALMAZÁSI LEHETŐSÉGEI A TANKÖNYVELEMZÉSben

Sramó András
PTE, KTK

Kulcsszavak: tankönyvelemzés, szövegvizualizáció, többdimenziós skálázás

A múlt század végének informatikai fejlődése a vizualizáció terén is új lehetőségeket teremtett nagy mennyiségű szöveges adatok kezelésére. Szövegvizualizációs technikák alkalmazása különösen akkor indokolt, ha nagy mennyiségű, ám alacsony strukturáltságú, dokumentum(ok) formájában rendelkezésre álló szöveges adatok belső szerkezetét és természetét kívánjuk megismerni, feltárni. Ezen technikák szokásos közelítésmódja, hogy a szöveget először egy magas dimenziószámú adathalmazba konvertálják, majd többváltozós statisztikai módszerek alkalmazásával az így előállított adathalmaz alacsonyabb, néhány dimenziós vetületeit szerkesztik meg, amelyek már grafikusan is megjeleníthetők. Néhány alkalmazás betekintést nyújt szövegszerű adathalmazokba, feltárva dokumentumok hasonlósági struktúráját (pl. WebSOM, ThemeSpace), adott szöveg időbeli menetét, adott téma fejlődését (ThemeRiver), illetve a gyakoribb szavak és kifejezések kapcsolatait (Pajek).

A szövegvizualizáció egy szöveg(csoport) matematikai szerkezetét kívánja feltárni. Ennek a folyamatnak három szintjét különböztethetjük meg:

- elsőrendű statisztikák alkalmazása: szóstatisztikák, gyakoriságok, konkordanciák;
- másodrendű statisztikák: klaszterező algoritmusok a dokumentumon belüli vagy dokumentumok közötti szerkezet feltárására;
- harmadrendű statisztikák: szemantikai szerkezet feltárása mesterséges intelligencia algoritmusokkal.

Ezeket az eljárásokat hol szövegelemzésnek, hol tartalomelemzésnek nevezik, és elmondható, hogy jól bejáratott, ma már döntő többségükben automatizált eszközök állnak rendelkezésünkre az elemzés végrehajtásához. A legtöbb szoftver a másodrendű statisztikák alkalmazásáig jut el következő lépéseken keresztül:

- A szövegbázis (korpusz) rögzítése. Ekkor még nincsenek kulcsszavak, témák, témakörök, előre definiált struktúrák sem a vizualizációhoz, sem az elemzéshez.
- Tartalomelemzés végrehajtása. A megfogalmazott elemzési szempontok alapján valamilyen automatizált eszközzel statisztikai feldolgozásra alkalmas adatok előállítása. Tankönyvek esetén nincs szükség teljes szövegfeldolgozásra, hanem egy konkordancia alapján kulcsszavakat célszerű kiválasztani, és ezek abszolút és relatív elhelyezkedését vizsgálni a szövegben.
- Klaszterelemzés végrehajtása. Ezzel a lépéssel vázolható fel a vizsgált szöveg belső struktúrája a megadott elemzési szempontok szerint.
- Többdimenziós skálázás. A klasztercentroidok elhelyezése kétdimenziós térben.
- Az eredmény grafikus megjelenítése.

Az előadás egy kiválasztott tankönyv példáján mutatja be a fenti elemzési folyamatot, és animációk segítségével illusztrálja a kapott eredmények többcélú hasznosíthatóságát. A tankönyvelemzés módszertani lehetőségeinek statisztikai és számítógépes megoldási lehetőségeit a pedagógiai praxis szempontjából a digitális tananyagok terjedése különösen is indokolttá teszi.