**Vízgazdálkodási adatok szemléletformáló adat-vizualizációja az optimumtól való eltérés alapján**

**Pitlik Marcell**

*12. évfolyam, Premontrei Szent Norbert Gimnázium, Gödöllő*

*Felkészítő tanár: Dr. Seres István, My-X kutatócsoport*

A My-X kutatócsoport tagjaként már részt vettem az Excel Solver funkciójával is hatékonyan megoldható problémakörök feltárásában. Ennek a tovább gondolásaként kezdtem el robottanárt fejleszteni, aminek az első lépése egy geometriai számológép megalkotása volt, amit úgy kell elképzelni, mint egy hagyományos számológépet sok-sok gombbal, csak a gombokkal síkgeometriai alakzatok paramétereit lehet megadni, ami alapján a geometriai számológép minden más adatot kiszámol. Ezen kívül a kutatócsoport által kifejlesztett hasonlóságelemzési eljárás felhasználási területeivel foglalkoztam, mint például komplex ár/érték-arányok vizsgálatával számítógépek esetében.

Az idén elhunyt Hans Rosling emlékére, aki az adat-vizualizáció nagymestere, kezdtünk el foglalkozni mi is az adat-vizualizációban rejlő lehetőségekkel. Célunk a hasonlóságelemzés és egyéb adatbányászati eljárások felhasználásával egy adatvagyon komplexebb, de közérthető bemutatása.

Az elterjedt statisztikai mutatók, mint pl. egy főre, egy férfira/nőre, egy négyzetkilométerre jutó természeti erőforrások, ill. pl. a GDP mellett, de ezek felhasználásával készítünk egy új mutatót is hasonlóságelemzéssel, ami mindezeket figyelembe véve felállít egy új sorrendet, az alapján, hogy az összes mutató figyelembe véve, melyik országban mennyi az egy főre jutó összes erőforrás-potenciálja a többi országhoz (ill. dinamikusan ország-év objektumhoz) képest.

A Rosling-féle animációk pl. az időben haladva mutatják be, miként változott az egyes országok egy főre jutó GDP-je a gyermekhalálozási ráta függvényében, úgy hogy az országokat jelölő színes korongok mérete arányos az ország lakosságával: vö. pl. (<https://www.ted.com/talks/hans_rosling_shows_the_best_stats_you_ve_ever_seen#t-563201>)

A pályázat keretében az ingyenesen bárki számára elérhető vízgazdálkodási adatokat használtam az egész világra nézve: vö. (<http://www.fao.org/nr/water/aquastat/data/query/index.html>), ill. adat-vizualizációs keretrendszerként az ismét csak közhasznú GAPMINDER program került adaptálásra.

A nyers adatok 5 éves lebontásban napjainktól 1960-ig állnak rendelkezésre országonként és régiónként is, ami 8 idő-fázist jelent. Mivel a nyers adatok sosem tartalmaznak teljes adattáblát, így az első feladat egy kisebb, de már teljes adattábla meghatározása, ha ez túl kicsi lenne, akkor az adathiányokat lehet közelítő módszerrel pótolni. pl. a két szomszédos érték átlaga, ha a jelenség linearitást mutat. Fontos az adatok ellenőrzése is, már amennyire ez lehetséges: pl. megegyezik-e a teljes afrikai GDP az egyes afrikai országok össz. GDP-jével. A további elemzéssel létrehozott rangsorokhoz szükséges minden abszolút számból viszonyszámot generálni: pl. GDP-ből egy főre jutó GDP. A folyamat vége pedig az adatok vizuális megjelenítése, és az ebből következő tények meghatározása.

A kutató csoporttal egy algoritmus fejlesztésén is dolgozunk, ami a nyers adattáblából automatikusan kiválasztja a legnagyobb teljes adattáblát, ha van ilyen, ha nincs akkor pedig a legkülönbözőbb adatfeltöltési eljárások (ha egy adat hiányzik csak, függvényt illeszt az adathalmazra és leolvassa a hiányzó értéket, ha az eltérés négyzete nem haladja meg egy megadott küszöbértéket) és így definiálja az algoritmus a rész/egész adattáblát.