



A szövegbányászat megjelenése a könyvtár- és információtudományi szakirodalomban

A szövegbányászat egyike azoknak a gyorsan fejlődő tudományterületeknek, amelyek segítik az információ kinyerését rendezetlen adathalmazokból. A terület multidiszciplináris módszereit egyaránt alkalmazzák a nyelvészet, a könyvtár- és információtudomány, a számítástudomány, a műszaki tudományok és az élettudomány kutatói. Könyvtári szakemberek elsősorban bibliometriai kutatásaikhoz hasznosítják a szövegbányászat eszköztárát. A Web of Science adatbázis alapján 1999 és 2013 között több mint 36 ezer tanulmány foglalkozott szövegbányászattal. Ezen belül a legtöbb írás az informatika, mérnökinformatika, bioinformatika, biokémia területén született. A Web of Science-ben a megadott időszakban 987 szövegbányászati tanulmány található a könyvtár- és információtudomány területén. Az adatbázis alapján a szövegbányászat alakulásának időbeli áttekintése mellett meghatározhatók a témakör legfontosabb folyóiratai, kutatóintézményei és szerzői.

A tudományometriai jellemzőknek megfelelően a szövegbányászati szakirodalom is folyamatosan növekszik. A leginkább hivatkozott időszak a 2003 és 2009 között megjelent tanulmányoké; az ennél frissebb szakirodalomra természetesen még nagyobb mértékben várható további hivatkozások megjelenése. A szakirodalom legfőbb kibocsátó országa az Amerikai Egyesült Államok, ahonnan a tanulmányok 40%-a származik. Emellett az összes hivatkozás több mint fele amerikai szerzők írásaira hivatkozik. A tanulmányok negyede négy ázsiai országból származik (Kína, Tajvan, Dél-Korea, Japán), amelyekre az összes hivatkozások ötöde irányul. Az egy cikkre eső legmagasabb hivatkozásszámot az amerikai, brit és ausztrál szerzők írásai érték el: az ő esetükben tíz fölött van az egy-egy tanulmányra érkező hivatkozások átlagos száma. Az amerikai szerzők elsősorban orosz, török, svájci és kanadai kutatókkal működtek együtt a témában; magyar szerzők pedig francia,

belga, osztrák, holland, német és kínai kutatókkal jegyeznek közös cikkeket.

A könyvtár- és információtudományi szakirodalomban található szövegbányászati tanulmányok 71 folyóiratban jelentek meg. Ezek közül három folyóirat közölte a tanulmányok több mint felét (*Journal of American Medical Informatics Association*, *Information Processing & Management*, *Journal of the American Society for Information Science and Technology*). A három folyóirat tanulmányaira érkezett a hivatkozások 61%-a. Az egy cikkre jutó legmagasabb hivatkozásszámot az *Information & Management* (átlagosan 24 hivatkozás egy cikkre), az *Information Systems Research* (átlagosan 16 hivatkozás egy cikkre) és az *MIS Quarterly* (átlagosan 15 hivatkozás egy cikkre) folyóiratok érték el.

A tanulmányok ötödét tíz intézmény kutatói jegyzik. A három legtöbb írást megjelentető intézmény a Columbia University, a University Arizona és a Nanyang Technological University. Az egyes intézményeken belül általában az élettudományi részlegek munkatársai jegyzik a tanulmányokat. A legtöbb hivatkozás a Columbia University, a University of Utah és a Harvard University munkatársaira érkezett. A területen a legtöbb, 15-15 tanulmányt H.C.Chen és C. Friedman jegyzi, a legtöbb hivatkozás pedig G. Hripcsak és C. Friedman írásaira érkezett.

A Web of Science adatai alapján a fentiekben azonosíthatók a könyvtár- és információtudományon belüli szövegbányászati kutatások magfolyóiratai, legfontosabb kutatóintézetei és szerzői.

/Shubhada Prashant Nagarkar, Rajendra Kumbhar: Text mining: An analysis of research published under the subject category 'Information Science Library Science' in Web of Science Database during 1999-2013. = LibraryReview, 64. köt. 3. sz. 2015. p. 248-262./

(Lencsés Ákos)