

Különböző folyóirattípusok automatikus statisztikai osztályozása

A probléma meghatározása

Ismeretes, hogy minden tudományágat kialakulásának kezdetén sommásan rendszerbe foglalnak. Ennek során a kvantitatív, főleg *statisztikai módszerek* használata figyelhető meg.

Jelenleg egy dokumentum helyének megállapítása és azonosítása valamely típus határain belül lényegi szinten történik. Megfigyelhető, hogy a cél típusformáló attribútumként jelenik meg a könyv esetében: meghatározza beállítottságát, stílusát, nyelvét, külalakját, és valamilyen módon még a tartalmát is. Feltehető, hogy a cél a folyóiratoknál is elsődleges tényező, és meghatározza a folyóiratok differenciálódását (tudományos, kutatási-termelési, termelési vagy műszaki folyóiratok stb.). Ha ez beigazolódik, akkor a folyóirattípusok statisztikai paraméterekkel is megkülönböztethetők.

Nyelvstatisztikai vizsgálatok mutatják, hogy a statisztikai módszerek alkalmasak a beszédstílusok megkülönböztetésére. Meghatározó attribútum a szóalak aritmetikai jelentése, hossza és eloszlása. Ezen alapult a cíkcímek hosszának mint a folyóirattípusok attribútumának vizsgálata a mezőgazdasági folyóiratok esetében. Jelenleg nincs egyértelmű mennyiségi kritérium a folyóiratok osztályozására; az egyetlen *Garfield* javaslata, miszerint egy tudományos folyóiratot az irodalomjegyzék és az évenként közreadott cikkek száma (minimum 20) jellemez.

Az 1970-es években megjelent mezőgazdasági folyóiratok osztályozása

A mezőgazdasági folyóiratok kiválasztása, a kiindulói adatok megállapítása és megjelenítése

A vizsgálat két szakaszban folyt. Külön vizsgálták az 1970-es évek folyóiratait, amikor a folyóiratok statisztikai osztályozásának a gondolata egyáltalán felmerült. Az ebben az időben végzett válogatást, és a Mir számítógépen folytatott adatfeldolgozást az 1980-as évek mezőgazdasági folyóiratainak Microsoft Excel programra támaszkodó feldolgozása követte. Ezért az 1970-es évek folyóiratainál a kiinduló adatok meghatározását és megjelenítését a korábbi technológia alapján végezték.

Első lépésként a mezőgazdasági folyóiratok három csoportját (*tudományos, kutatással és termeléssel, valamint termeléssel foglalkozók*) választották ki a címlapon szereplő meghatározás alapján.

Ez a Szovjetunióban az adott időszakban megjelent mezőgazdasági folyóiratok 54,2%-át jelentette. Ezt követően megvizsgálták a szavak eloszlását az adott folyóirattípusban megjelent cikkek címében. A grafikus feldolgozás lehetővé tette, hogy képet alkothassanak az egyes folyóirattípusokról, valamint egy adott típushoz tartozó folyóiratokról.

A kiválasztott folyóiratok tartalmi vizsgálata alapján világossá vált, hogy azok, amelyeknek a címe „*Vestnik*” (Közlöny, Közlemények) szóval kezdődik, a tisztán tudományos problémák mellett a tudományos kutatások szervezésével stb. is foglalkoznak. A kiindulási adatok eloszlásának elemzése azt mutatta, hogy ez a célkitűzés befolyásolhatja az ilyen folyóiratokban megjelenő cikkek címeinek hosszát, mivel ezeknek a folyóiratoknak a terjesztése eléggé különbözik a tudományos folyóiratokétól. Ennek alapján a kiválasztott folyóiratokat további két csoportra osztották: *tudományos folyóiratokra* (4809 cikk) és *tudományos-szervezési folyóiratokra* (2013 cikk).

A kutatási-termelési folyóiratok cikkeinél a címek szóeloszlása számos esetben nagyon hasonlít a termelési folyóiratok eloszlásához, ezért ezeket érdemes az utóbbi csoportba sorolni. A kutatási-termelési folyóiratok 7360 cíkcímét képviseltek, a termelési folyóiratokban 8450 cíkcím volt.

A mezőgazdasági folyóiratok statisztikai osztályozása

Az empirikus szóeloszlás a kiválasztott folyóiratok cíkcímeiben és a kialakított szakcsoportokban logaritmikus. Ez lehetővé tette, hogy a mezőgazdasági folyóiratokat a súlyozott számtani átlag (*várható érték*) és *módusz* alapján csoportosítsák.

Látható volt, hogy minden folyóirattípusnak saját intervalluma van, és megfelel a lényegi szinten végzett osztályozásnak. Érdemes megjegyezni, hogy a módusz nagy szemantikus erővel tűnik ki (ez jellemző az egyetlen móduszú logaritmikus eloszlásra).

A grafikus megjelenítés a vizsgált folyóiratok minőségi értékelését is segíti. A kérdés az, milyen várható érték és módusz jellemzi az egyes folyóirattípusokat. Ismeretes, hogy a varianciaelemzés nemcsak a normál, hanem más eloszlásokra is alkalmazható, így az egyes folyóirattípusokban

megjelent cikkek címeiben szereplő szavak várható számára és móduszára is elvégezhető. Az elemzés kimutatta, hogy a várható érték és a módusz a mezőgazdasági folyóirat típusok olyan mutatószámai, amelyek alapján a folyóiratok osztályozhatók.

Ezeknek a mutatószámoknak ismeretelméleti jelentőségük is van. Feltételezhető, hogy az adott jelenség az önszerveződés képét mutatja, amikor egy alacsonyabb szinthez tartozó paraméter egy magasabb szint viselkedésének a mutatószáma.

Az 1980-as években megjelent mezőgazdasági folyóiratok osztályozása

Az ebben az időszakban megjelent mezőgazdasági tárgyú folyóiratoknak már 90%-a tartozott a fenti csoportokba. Megállapították, hogy az 1970-ben elemzett folyóiratok mutatószámai lényegesen nem változtak. Az automatikus osztályozás szerint csak egy kutatási-termelési tárgyú folyóirat lépett át a tudományos-szervezési kategóriába, és három termelési folyóirat vált kutatási-termelési folyóirattá.

Az 1980-as minták móduszainak és várható értékeinek ábrázolására a poláris koordináta-rendszert alkalmazták. Ezeknek a paramétereknek a vizsgálata azt jelezte, hogy a folyóiratok mind a négy csoportja a statisztikai jellemzőkkel határolt térben foglalt helyet. Ennek alapján megkísérelték a folyóiratok *automatikus osztályozását* is elvégezni.

A kapott adatok alapján az automatikus folyóirat-osztályozási technológiát minden szakterületre kidolgozták. Ez három szakaszban történt:

1. szakasz: a folyóirat-osztályozás módusz, várható érték és varianciaelemzés alapján történő alkalmazása;
 2. szakasz: adaptálás bizonyos ismeretterületeken a folyóirat típusok mutatószámainak megállapítására;
 3. szakasz: az egyes folyóiratok besorolása.
- A 2. szakasz lépései:
- a) a cikkek címében szereplő szavak számának betáplálása;
 - b) adatmegjelenítés a folyóiratcsoport megállapítása céljából a következő módon:
 - a folyóiratminta jellemzése a cíkcímekben lévő szavak számának gyakoriságával (százalékban kifejezve);
 - az egyes folyóiratok és folyóiratcsoportok grafikus ábrázolása a cíkcímekben lévő szavak számának alapján (százalékban), egy vagy két kritérium egyidejű alkalmazásával;
 - c) a módusz és a várható érték megjelenítése minden folyóirat számára, vagy egyszerűen a helyük kijelölésével a szavak számának megfelelő tengelyen, vagy poláris koordináták segit-

ségével. Megjegyzendő, hogy az automatikus folyóirat-osztályozás javítása a statisztikai mutatószámok adott folyóirattípusra jellemző háttérértékének finomításával történhet. Ezt a feladatot valószínűleg a *Bayes-kritérium* alapján lehet megoldani.

Az automatikus statisztikai osztályozás lehetőségei

A további kutatások iránya az automatikus osztályozási technológia javítása mellett más statisztikai mutatók (pl. hivatkozások) alkalmazása lehet. A cél egy komplex folyóirat-osztályozási módszer kidolgozása, amely a könyvtárakban a dokumentumok kiválasztását, a pénzügyi támogatást, a hipertext rendszereket, a kivonatolást stb. pozitív irányban befolyásolhatja. A tisztán lingvisztikai mutatók alkalmazása is segítheti ezt a törekvést.

A statisztikai mutatók egyéb téren is hasznosak lehetnek, pl. segíthetik a folyóiratok referálásra történő kiválasztását, vagy a társadalomtudományi folyóiratok tipológiájának tanulmányozását. A *Science Citation Indexben* és a *Social Science Citation Indexben* található folyóiratok vizsgálata választhat arra, hogy vajon az összes, az Institute for Scientific Information (USA) által kiválasztott folyóirat tudományos, tudományos-szervezési vagy kutatási-termelési irányú-e.

Azt találták, hogy a folyóiratok tipológiai jellemzése lehetővé teszi a *Breadford-törvény* új értelmezését a folyóiratoknak a magban vagy az elosztás más zónáiban való elhelyezkedését illetően. Az ismeretek bizonyos területein, így a nemzetgazdaságban is alkalmas lehet a belső információ fejlődésének vizsgálatára.

Az automatikus folyóirat-osztályozás segítségével a folyóiratok fejlődése is nyomon követhető. A kapott jellemzőket nemcsak a szerkesztőbizottságok tagjai, a szerzők és a szerkesztők, hanem a tudománymetria és a bibliometria kutatói, valamint a különböző kiadványok összeállítói is haszonnal alkalmazhatják.

Megjegyzés: az eredeti közlemény számos táblázatban és ábrában teszi közzé a vizsgálat adatait és eredményeit.

Irodalom

- Letopis' periodičeskih i prodolžaŭsihsâ izdanij. 1986–1990. 1. rész – Moszkva, Knižnaâ palata, 1992. 207 p.
- Letopis' periodičeskih i prodolžaŭsihsâ izdanij. 1986–1990. 1. rész – Moszkva, Knižnaâ palata, 1988. 158 p.
- PIOTROVSKIJ R. T.–BEKTAEV K. B.–PIOTROVSKAĀ A. A.: *Matematičeskaâ lingvistika*. Moszkva, Vysšaa škola, 1977. 383 p.

UZILEVSKIJ G. Á.: Ispol'zovanie količestvennyh metodov v sistematike pervičnyh dokumentov (issledovanie žurnalov po sel'skomu hozjajstvu). = Naučnaâ tehničeskâ informaciâ, ser. 1. 5. sz. 1985. p. 22–26.

ČERNÁK A. A.: Istorijâ tehničeskoj. Moszkva, Kniga, 1981. 318 p.

GARFIELD, E.: Citation analysis as a tool in journal evaluation / E. Garfield. = Science, 178. köt. 4060. sz. 1971. p. 471–475.

/GILYAREVSKI, R.–UZILEVSKY, G.–MOUDROV, E.: An automatic statistical classification of different types of journals. = International Forum on Information and Documentation, 22. köt. 3. sz. 1997. p. 24–37./

(Zsindely Sándor)

Lengyel egyetemi könyvtárak folyóirat-adatbázisa

Kezdetben négy, napjainkban pedig kilenc központi egyetemi könyvtár alakított konzorciumot annak érdekében, hogy központi folyóirat-adatbázist hozzon létre. Az elhatározás valóra váltásának előfeltétele a VTLS integrált könyvtári rendszer megvásárlása volt.

Az 1993 és 1994 folyamán megtervezett adatbázist 1995 tavaszán kezdték feltölteni. 1997 közepére összesen 4289 rekordot halmoztak fel benne a részt vevő könyvtárak. Maga a bázis a gdański Központi Egyetemi Könyvtár Hewlett-Packard 3000 clas 02 számítógépére van telepítve.

A konzorcium által birtokolt folyóiratok bibliográfiai leírásai megfelelnek a szóban forgó kiadványfajtára kimunkált USMARC formátum követelményeinek. Míg a leírások természetesen a részt vevő könyvtárakban decentralizáltan készülnek, addig az „elválogagos” ellenőrzés a gdański Központi Egyetemi Könyvtárban történik.

Az adatbázis hozzáférhetősége három szinten van biztosítva, úm.

- a helyi hálózat szintjén (LAN = Local Area Network),
- városi szinten (MAN = Metropolitan Area Network),
- az Internet hálózatán (a címe itt: *telnet :cala.bg.univ.gda.pl*, parancs: *Hello user.clas02*).

A hozzáférést többféle keresőkulcs (pl. cím, kiadó intézmény, ISSN-szám) biztosítja. Az egyes címek közötti összefüggések is kimutathatóak az adatbázis használata közben.

Általános tapasztalat: minél több a részt vevő intézmény, annál több probléma fordulhat elő az adatbázis kurrens feltöltésében. Más kérdés: eddig sikerült megőrizni a rekordok egyértelműen magas színvonalát.

/SUSZKO-SOBINA, I.: Centralna baza czasopism (CKTCZ). = Bibliotekarz, 7–8. sz. 1997. p. 10–12./

(Futala Tibor)

A „kiegési” szindróma és a könyvtárosság

A „kiegés” szindrómájáról kb. másfél évtizede beszélnek. Definíciója: „Fizikai és pszichikai kimerültség, amelyet a saját lehetőségek végletes kizsákmányolása okozott.” Kísérő tünetei: apátia, reménytelenség, állandó elégedetlenség. Az általa érintettek fizikai szempontból mindig fáradtak, nehézségeik vannak a reggeli felkeléssel és az esti elalvással. Csökken a figyelmük, koncentrációs képességük, problémamegoldási készségük. Nehézségeik lehetnek emlékezetükkel is.

Közvetlen kiváltó oka a munkahely és a végzett munka jellege. Megjelenési formái felettébb változatosak. Úm.:

- a bizonyos mértékű függetlenség és önállóság hiányával párosuló bürokrácia;

- a megszakíthatatlan közönségkapcsolatok, a folyvást „első vonalban” végzett munka;
- egyenlőtlen megterhelés, a munkatársakra és személyiségekre érzéketlen feladatmeghatározás;
- inadekvát felügyelet, vagy több főnöknek való alárendeltség;
- a határozott célok és feladatok hiánya, az intézmény céljainak meghatározásából, a saját feladatok eldöntéséből való kizártság;
- az önmegvalósítás „alig lehetőségei” a végzett munkában;
- a munkavégzéssel kapcsolatos – főként pozitív – visszacsatolás hiánya;
- az ellenőrzés hiánya, vagy az ellenőrzés hiányának érzete;