

LBD AKO NOVÁ KOMPETENCIA INFORMAČNÝCH ŠPECIALISTOV V KONTEXTE ZNALOSTNEJ SPOLOČNOSTI

doc. PhDr. Marcela Katuščáková, PhD.; marcela.katuscakova@fses.uniba.sk; (Ústav mediamatiky, Fakulta sociálnych a ekonomických vied, Univerzita Komenského v Bratislave)

Vývoj nových informačných technológií spôsobuje výrazné sociálno-ekonomické zmeny. Nárast objemu dát a informácií v digitálnej forme presúva náš život do digitálneho informačného prostredia a zároveň zvyrazňuje potrebu nových kompetencií. Mení sa infraštruktúra vedy ako hlavného predmetu záujmu informačných vedcov. Potrebu nových služieb s pridanou hodnotou pre vedeckovýskumných pracovníkov môžu pomôcť naplniť systémy LBD (Literature-Based Discovery), zamerané na objavovanie implicitných vzťahov vo vedeckej literatúre.

<http://doi.org/10.52036/1335793X.2022.SC.5-9>

Dlhodobý trend vývoja a zjednodušenia prístupnosti (sieťových) informačných technológií výrazne prispel k celosvetovým socio-ekonomickým zmenám. Pozorujeme postupnú difúziu nových informačných priestorov do života spoločnosti smerom k životu v digitálnom resp. hybridnom informačnom priestore (Konvit et al. 2019). Prístup k informáciám sa zjednodušuje, následkom čoho rastie význam kognitívnych zručností, potrebných na premenu veľkých objemov dát a informácií na využiteľné znalosti. Pozornosť odborníkov sa zameriava ako na organizovanie obrovských objemov dát a informácií, tak aj na možnosti objavovania využiteľných znalostí z nich. Informačná spoločnosť, zameraná primárne na nové informačné technológie, vytvorila vhodné podmienky na transformáciu informácií na znalosti a ich následnú aplikáciu do praxe. Predstavuje významný predpoklad nástupu znalostnej spoločnosti, nie je však automatickou garanciou tohto posunu. Bez potrebných kompetencií sa naplnenie predstáv znalostnej spoločnosti nemusí v dostatočnej miere zrealizovať. Sledujeme zmeny v požiadavkách trhu práce a knižnično-informačná veda (KIV) čelí novým výzvam aj hrozbám. Pokúša sa identifikovať nové zručnosti potrebné pre 21. storočie (Norlander a Barchas-Lichtenstein, 2021) a zároveň diskutuje o možnostiach rozšírenia oblastí jej poznania (Capurro a Zins 2017). Informačná veda (IV) ako interdisciplinárny vedný odbor sa od svojho začiatku zameriava na riešenie problémov v kontexte nárastu vedecko-technických informácií (Bawden, Robinson 2012). Nakoľko veda a inovácie sa stávajú hlavným indikátorom úrovne zna-

lostnej spoločnosti, je úloha IV pre spoločnosť stále významnejšia. Navyše, ako sme už konštatovali v úvode, mení sa informačná a znalostná infraštruktúra vo vede a veľká časť vedeckého procesu sa presúva do digitálneho informačného prostredia. Digitálna a otvorená veda sa tak stáva predmetom skúmania informačných vedcov, ktorí sa pokúšajú vybudovať nové modely služieb s pridanou hodnotou pre vedeckovýskumných pracovníkov (Steinerová 2016). Aktuálne informačné nástroje, ktoré napomáhajú vedeckým pracovníkom sú založené na tradičnom prístupe IV, zameranom na vyhľadanie explicitných informácií a explicitných bibliografických vzťahov v rámci databáz vedeckých textov. Ide o prácu s objavenými a publikovanými poznatkami. Otázkou je, či má IV možnosti posunúť sa vo svojej snahe pomáhať vede a vedeckým pracovníkom ešte ďalej. Don R. Swanson bol presvedčený, že tento potenciál tu je, pokiaľ sa zručnosti informačných pracovníkov posunú o krok ďalej. Ešte na konci 80. rokov minulého storočia predstavil metódu LBD (Literature-Based Discovery) ako rozšírenie procesu informačného prieskumu smerom k objavovaniu implicitných prepojení v rámci databáz vedeckých textov. Dnes v kontexte budovania znalostnej spoločnosti cez podporu vedy a inovácií možno metódu LBD vnímať ako výzvu na ďalšie rozšírenie kompetencií informačných profesionálov v prostredí digitálnej a otvorenej vedy.

NEOBJAVENÉ VEREJNE PRÍSTUPNÉ POZNATKY

Prof. Don R. Swanson, dekan Graduate Library School na University of Chicago, si všimol, že poznatky môžu

byť verejne známe, teda publikované, a v rovnakom čase neznáme alebo neobjavené. Znalosti sú často roztrúsené v množstvách rôznych vedeckých časopisov a ich zmysluplné prepojenie je často otázkou náhody, ako tomu bolo aj v jeho prípade. Swanson náhodne postrehol, že niekoľko príznakov, ktoré sa vyskytujú o pacientov s tzv. Raynaudovým syndrómom, ktorým trpel jeho známy, korešponduje so zmenami vyvolanými pozitívnym rybieho oleja, ale v opačnom zmysle (o vplyve rybieho oleja na zdravie si náhodne prečítal v popularizačnom článku o Eskimákoch žijúcich v Grónsku). Z toho mu vyplynulo, že rybí olej by sa mohol používať na potlačenie niektorých symptómov Raynaudovho syndrómu (Smalheiser 2017). Ako informačný pracovník, pred tým, než svoje odporúčanie komunikoval ďalej, sa rozhodol potvrdiť si toto prepojenie v databáze biomedicínskych článkov. Zistil však, že množiny literatúry o Raynaudovom syndróme (asi 2000 článkov) a rybom oleji (asi 1000 článkov) boli neprepojené, teda že v danom čase (1986) neexistoval vedecký článok, ktorý by tieto dve oblasti logicky prepájal. Všimol si, že tieto dve množiny mali mnohé pojmy spoločné, napr. viskozita krvi, agregácia krvných doštičiek a pod. Rozhodol sa pokúsiť nájsť bibliografické prepojenie formou citácií, avšak zistil, že žiaden článok z prvej množiny literatúry neodkazuje na žiaden článok z druhej množiny literatúry, vrátane kocitácií autorov z oboch množín literatúry. Všetko poukazyvalo na to, že išlo o dve izolované množiny literatúry (disjunktné). Následne zrealizoval obsahovú analýzu oboch množín literatúry a hľadajúc ich logické prepojenie excerpoval často sa opakujúce významné pojmy z oboch množín. Výsledky obsahovej analýzy sa pokúsil zaznačiť formálnym zápisom:

Ak

- A = rybí olej,
- B = zníženie agregácie krvných doštičiek
- C = zlepšenie Raynaudovho syndrómu

potom, môžeme konštatovať, že: A „spôsobuje“ B, B „spôsobuje“ C. Z toho mu vyplynula hypotéza: A „spôsobuje“ C (Swanson 1986). Zaujímavé je, že táto hypotéza vytvorená informačným pracovníkom len na základe informácií z vedeckých databáz sa následne v klinických biomedicínskych štúdiách potvrdila (Chang et al. 1988; DiGiacomo et al. 1989).

O dva roky neskôr Swanson prišiel s novou hypotézou, že magnézium (biologicky aktívna látka) môže liečiť príznaky migrény (ochorenie). Postupoval podobne ako v prvom prípade, identifikoval dve množiny vedeckej literatúry a zisťoval, či ide o neprepojené množiny. Ani v tomto prípade nenašiel bibliografické prepojenie

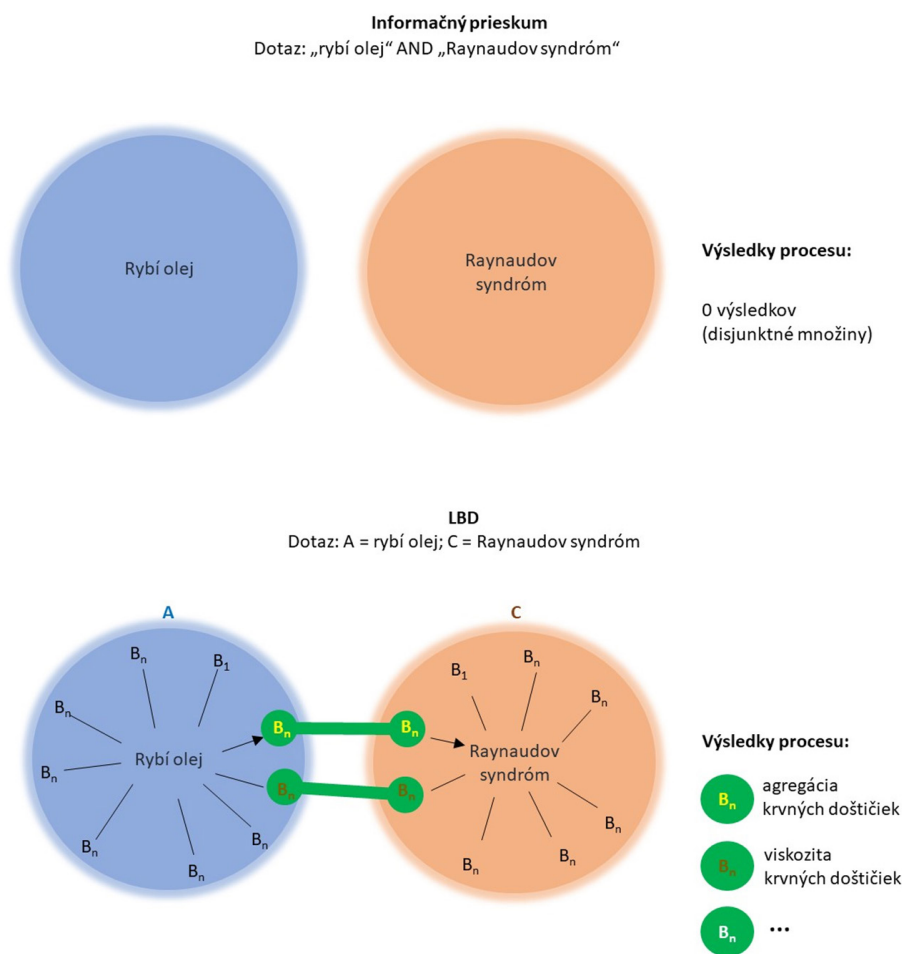
medzi množinami literatúry o migréne a o magnéziu. Na základe podobného postupu ako v prvom prípade našiel až 11 logických prepojení (napr. cez stres, vaskulárny tonus, blokátor kalciového kanála), ktoré podporovali jeho hypotézu, že magnézium môže pomáhať pacientom s migrénou (Swanson 1988). Na základe článku publikovanom v lekárskom časopise, kde predstavil svoju hypotézu podloženú logickým usudzovaním, bola jeho hypotéza experimentálne potvrdená (Castelli et al. 1993; Facchinetti et al. 1991; Gallai et al. 1993; Thomas et al. 1994)“plainCitation“:(Castelli et al. 1993; Facchinetti et al. 1991; Gallai et al. 1993; Thomas et al. 1994.

Pokúsil sa tak navrhnuť všeobecný model objavovania komplementárnych doposiaľ disjunktných množín, ktoré nemajú žiaden spoločný článok, navzájom sa necitujú, neodkazujú na seba a nie sú spoločne kocitované. Uvedomoval si, že pokiaľ sú dve množiny literatúry prepojené pomocou logických argumentov, ktoré pri-nášajú, potom očakáva, že medzi nimi bude existovať aj nejaké bibliografické prepojenie. Pokiaľ sa tam nenachádza, v tom prípade by sa malo stať predmetom záujmu vedcov.

ABC – MODEL OBJAVOVANIA

Swanson vypracoval postup objavovania implicitných prepojení disjunktných množín vedeckej literatúry, kde pracuje s dvoma množinami literatúry: A (cieľová množina) a C (počiatočná), prepojené cez množinu B (stredných termínov). Neskôr bolo toto poradie upravené do abecedného poradia A (počiatočná), B (stredná), C (cieľová) množina. Swanson pri svojej práci postupoval nasledovne: pokiaľ sa v rámci jednej množiny literatúry uvádza, že A spôsobuje B (napr. liečivo A spôsobuje zmenu B v krvi), potom takéto tvrdenie označil AB. Pokiaľ sa v rámci druhej množiny literatúry uvádza, že B spôsobuje C (zmena B v krvi má vplyv na chorobu C), potom toto tvrdenie označil BC. Predpokladal, že pri pohľade na takéto zápis by viacerým napadlo uvažovať na možnosťou vzťahu medzi A a C ako AC. Táto štruktúra modelu ABC pripomína sylogizmus, avšak Swanson upozorňuje, že medzi termínmi A, B, C sa využíva vzťah asociácie, ktorý nie je tranzitívnym vzťahom, a preto sa musí prirovnanie k sylogizmu vnímať len metaforicky (Swanson 2008).

Pokiaľ by sme chceli porozumieť rozdielom medzi procesom tradičného informačného prieskumu a zatvoreného prístupu LBD pri vyhľadani dvoch disjunktných množín v spomínanom príklade rybieho oleja a Raynaudovho syndrómu, potom v prípade využívaných postupov vyhľadávania v databázach vedeckých článkov, Swansonovi (v roku 1988, keď sa ešte o prepoje-



Obr. 1 Porovnanie výsledkov tradičného informačného prieskumu a (zatvoreného) modelu LBD pri vyhľadávaní dvoch disjunktných množín literatúry A (rybí olej) a B (Raynaudov syndróm).

ní nevedelo) pri zadaní dotazu „rybí olej“ AND „Raynaudov syndróm“ systém ponúkol 0 výsledkov. Avšak s integráciou LBD do vyhľadávania vedeckej literatúry by dostal zoznam stredných B termínov, ktoré môžu tieto dve oblasti logicky prepájať a viesť tak k novým vedeckým hypotézam.

Swanson rozpracoval dva varianty LBD, tzv. otvorený a zatvorený model LBD. Otvorený/jednouzlový variant objavovania vychádza z predpokladu, že poznáme len počiatočnú oblasť objavovania, resp. termín A, pričom cieľom je nájsť možné prepojenia na neznáme cieľové termíny C cez jeden alebo viacero stredných termínov B. Hľadajú sa termíny A, B, C pričom vzťahy medzi A-B a B-C sú známe. Avšak vzťah medzi A-C nebol doposiaľ identifikovaný. Zatvorený/dvojuzlový variant vyžaduje, aby boli počiatočný termín A aj cieľový termín C preddefinované používateľom a cieľom LBD je nájsť

nový vzťah AB a BC, ktorý umožní prepojiť doposiaľ neprepojené témy A a C. Takže zatvorený variant sa zameriava na vyhľadanie stredných termínov B, otvorený variant vyhľadáva aj stredné B termíny a následne cieľové termíny C (Sebastian et al. 2017).

Kým v roku 1986 pri analýze publikácií o Raynaudovom syndróme a rybom oleji na základe vyhľadávania v databáze PubMed Swanson „manuálne“ analyzoval asi 3 000 článkov, v roku 1997 by to už bolo viac ako 10 000 a v roku 2019 asi 40 000. Aj z tohto dôvodu nebol o túto metódu na konci 20 storočia ešte veľký záujem, avšak s vývojom technológií sa možnosti automatizácie procesov LBD stále zlepšujú. Swanson si uvedomoval potrebu technologickej podpory procesu LBD, a tak spolu so svojim kolegom v roku 1997 uviedli prvý automatizovaný nástroj LBD pod názvom Arrowsmith (Swanson a Smalheiser 1997). Nástroj

pomáhal pri tých fázach LBD, pri ktorých boli funkcie tradičných prieskumových nástrojov nepostačujúce, napr. pri extrakcii významných slov alebo fráz spoločne sa vyskytujúcich v počítačovej a cieľovej množine literatúry či odstránení neplnovýznamových stop-slov, alebo zoradení zoznamu stredných B termínov podľa relatívnej frekvencie B termínov. Neskôr systém ponúkal možnosť výberu sémantických kategórií stredných termínov alebo nastavenie hodnoty dátumu prvého publikovania a pod. Nástroj Arrowsmith je možné využívať dodnes.

S vývojom poznania v oblasti dolovania znalostí z textov (textmining) a v oblasti znalostných systémov sa začali práce na ďalších systémoch LBD často s využitím nástrojov ako PubMed, MeSH, UMLS, SemMedDB, na vývoji ktorých pracuje Národná lekárska knižnica USA (NLM – National Library of Medicine). Prvé vyvinuté nástroje LBD vychádzali z princípov informačného prieskumu a zvyknú sa pomenovávať ako štatistické systémy kam patrí aj systém Arrowsmith. Neskôr sa vyvíjali LBD systémy založené na znalostiach, ako napr. systém BITOLA, DAD, Manjal, kde sa prešlo od práce s termínmi smerom k pojmom. Aktuálne sa vyvíjajú tzv. hybridné systémy LBD, ktoré v sebe integrujú prvky štatistických aj znalostných systémov, rovnako tak strojové učenie a vizualizačné nástroje. Vhodným príkladom je systém LION LBD vyvinutý na Univerzite v Cambridge.

PRAKTICKÁ APLIKÁCIA SYSTÉMOV LBD

Uvedené systémy majú svoju aplikačnú oblasť biomedicíny, kde je práca s jazykom formalizovaná a kde sa využívajú nástroje ako tezaurus MeSH, ontológia UMLS či sémantický nástroj Semantic MEDLINE. Pomocou metódy LBD sa podarilo identifikovať množstvo nových objavov, resp. vzťahov v oblasti biomedicíny, ako napr. vzťah medzi somatomedin C a arginine, nedostatok magnézia a neurologické ochorenia, indomethacin a Alzheimer, estrogén a Alzheimer, Calcium independent phospholipase A2 a schizofrénia. Zároveň sa podarilo objaviť: lekárske využitie kurkumínu, potenciálnu liečbu rakoviny, šedého zákalu, roztrúsenej sklerózy, rakoviny prsníka, rast srdcovej hypertrofiie, Parkinsonovej choroby či rakoviny prostaty. LBD sa využíva aj pri vývoji nových liekov, predpovedaní nežiaducich účinkov liekov, novom využití existujúcich liekov a pod.

V poslednej dobe sme svedkami prvých aplikácií modelov LBD mimo oblasti biomedicíny, napr. v rámci humanitných vied, kde sa nehľadajú vzťahy príčiny a následku, ale interpretácie alebo analógie. Problém v tomto kontexte je s automatickým spracovaním pri-

rodzeného jazyka (nízka formalizácia). V roku 1999 bola aplikovaná metóda LBD v databáze Wilson (EBSCO), kde sa hľadali neidentifikované prepojenia medzi básnikmi 20. storočia a antickými filozofmi (Cory 1999). V snahe predísť nejednoznačnostiam jazyka sa pracovalo s termínmi ako sú mená autorov alebo hlavné oblasti výskumu, pričom výskum priniesol nové zaujímavé prepojenia.

ZÁVER

Práca so systémami LBD ešte nie je úplne jednoduchá, čo potvrdzuje aj náš výskum spoločne s expertkou z oblasti biomedicíny prof. RNDr. Erikou Halašovou, PhD. Pri objavovaní sme narazili na viacero problémov ako napr. príliš veľký počet odporúčaných stredných B termínov. Je potrebné si uvedomiť, že práca so systémami LBD je poloautomatizovaná a stále si vyžaduje prácu informačného profesionála a odborníka z danej oblasti výskumu (Katuščáková 2020). Predpokladáme, že vývoj systémov LBD bude rýchlo postupovať, aj vzhľadom na pandémiu COVID-19, kedy v úvode pandémie boli publikované medicínske články, konštatujúce problémy s objasnením súvislostí prečo niektorí majú ťažší priebeh ochorenia ako iní a volali po zapojení odborníkov na LBD, ktorí by mohli tieto implicitné vzťahy vedieť identifikovať. Budeme radi, ak odborníci na informačnú vedu prejavia záujem o metódu LBD aj v našich zemepisných šírkach. Domnievame sa, že integrácia metódy LBD napr. v rámci vedeckých knižníc pre vedeckých pracovníkov z oblasti biomedicíny, by mohla byť veľmi prínosná a zároveň by sa využitím metódy pracujúcej so zaznamenaným poznaním mohli ušetriť nemalé finančné prostriedky spojené s biomedicínskym výskumom. Uvedomujeme si, že by išlo o posun do menej známych oblastí KIV. Metóda LBD však bola vytvorená odborníkom z informačnej vedy, ktorý ju vnímal ako rozšírenie informačného prieskumu. Nie je prekvapivé, že prvé miesta vedeckého záujmu o LBD aktuálne preberajú oblasti ako počítačová veda a lekárska informatika, avšak hneď za nimi sa ešte stále umiestňuje KIV. Považujeme za výzvu, pokúsiť sa v kontexte znalostnej spoločnosti, digitálnej a otvorenej vedy rozšíriť kompetencie informačných profesionálov smerom k objavovaniu implicitných vzťahov v rámci verejne prístupných vedeckých poznatkov.

■ Príspevok bol spracovaný v rámci riešenia projektu APVV-18-0479 Výskum kľúčových kompetencií pre znalostnú spoločnosť v kontexte historických, sociálnych a ekonomických špecifik SR.

■ článok je recenzovaný

Zdroje:

Zdroje:

BAWDEN, David a Lyn ROBINSON, 2012. *The Facet LIS Textbook Collection: Introduction to Information Science*. London: Facet Publishing. ISBN 978-1-85604-810-1.

CAPURRO, R. a Ch. ZINS, 2017. Knowledge map of information science [online]. 2017. Dostupné na: <http://www.capurro.de/zins.html>

CASTELLI, S., C. MEOSSI, R. DOMENICI, F. FONTANA a G. STEFANI, 1993. [Magnesium in the prophylaxis of primary headache and other periodic disorders in children]. *La Pediatria Medica E Chirurgica: Medical and Surgical Pediatrics*. 1993, roč. 15, č. 5, s. 481–488. ISSN 0391-5387.

CORY, K.A., 1999. Discovering hidden analogies in an online humanities database. *Library Trends*. 1999, roč. 48, č. 1, s. 60–71.

DIGIACOMO, R. A., J. M. KREMER a D. M. SHAH, 1989. Fish-oil dietary supplementation in patients with Raynaud's phenomenon: a double-blind, controlled prospective study. *The American Journal of Medicine*. 1989, roč. 86, č. 2, s. 158–164.

FACCHINETTI, F., G. SANCES, P. BORELLA, A. R. GENAZZANI a G. NAPPI, 1991. Magnesium prophylaxis of menstrual migraine: effects on intracellular magnesium. *Headache*. 1991, roč. 31, č. 5, s. 298–301. ISSN 0017-8748.

GALLAI, V., P. SARCHIELLI, P. MORUCCI a G. ABBRITTI, 1993. Red blood cell magnesium levels in migraine patients. *Cephalalgia: An International Journal of Headache* [online]. 1993, roč. 13, č. 2, s. 94–81; discussion 73. ISSN 0333-1024. Dostupné na: doi:10.1046/j.1468-2982.1993.1302094.x

CHANG, B. B., R. DIGIACOMO, J. KREMER, C. KAY a D.M. SHAH, 1988. Effects of fish oil fatty acid ingestion in patients with Raynaud's syndrome. *Surgical Forum*. 1988, roč. 39, s. 324–326.

CHEN, D. H. C. a C. J. DAHLMAN, 2006. *The Knowledge Economy, The Kam Methodology And World Bank Operations* [online]. Washington: World Bank Institute. Dostupné na: <https://documents1.worldbank.org/curated/en/695211468153873436/pdf/358670WBI0The11dge1Economy01PUBLIC1.pdf>

INSTITUTE OF MUSEUM AND LIBRARY SERVICES, no date. *Museums, Libraries, and 21st Century Skills | Institute of Museum and Library Services* [online] [cit. 20.5.2022]. Dostupné na: <https://www.ims.gov/issues/national-initiatives/museums-libraries-and-21st-century-skills>

KATUŠČÁKOVÁ, Marcela, 2020. *Od vyhľadávania informácií k objavovaniu nových znalostí: Swansonova metóda prepájania pojmov*. Bratislava: Univerzita Komenského v Bratislave. ISBN 978-80-223-5072-3.

KONVIT, M., E. CAPKOVÁ a J. GREČNÁR, 2019. *Virtuálne priestory*. Bratislava: Univerzita Komenského v Bratislave. ISBN 978-80-223-4858-4.

KOSTOFF, R.N., 2008. Literature-Related Discovery (LRD): Introduction and background. *Technological Forecasting & Social Change*. 2008, roč. 75, s. 165–185.

NORLANDER, R. J. a J. BARCHAS-LICHTENSTEIN, 2021. *Skills for 21st-Century Librarians: Learning Objectives for Library Programming* [online]. Chicago: American Library Association [cit. 12.5.2022]. Dostupné na: <https://www.ala.org/tools/sites/ala.org.tools/files/content/211022-ppo-skills-for-21st-century-librarians-report-SCREEN.pdf>

POPPER, Karl R., 1972. *Objective Knowledge: An Evolutionary Approach*. Revised edition. Oxford Eng.: New York: Oxford University Press. ISBN 978-0-19-875024-6.

SEBASTIAN, Yakub, Eu-gene SIEW a S. ORIMAYE, 2017. Emerging approaches in literature-based discovery: techniques and performance review. *The Knowledge Engineering Review*. 2017, roč. 32, č. 12, s. 1–35.

SMALHEISER, Neil R., 2017. Rediscovering Don Swanson: the Past, Present and Future of Literature-Based Discovery. *Journal of data and information science (Warsaw, Poland)* [online]. 2017, roč. 2, č. 4, s. 43–64. ISSN 2096-157X. Dostupné na: doi:10.1515/jdis-2017-0019

STEINEROVÁ, J., 2016. *Informačná veda: tradície, teórie a trendy - smerom k digitálnej vede*. V: *Knižničná a Informačná Veda: Zborník Filozofickej Fakulty Univerzity Komenského*. XXVI, s. 7–27. ISBN 978-80-223-4252-0.

SWANSON, Don R., 1986. Fish Oil, Raynaud's Syndrome, and Undiscovered Public Knowledge. *Perspectives in Biology and Medicine*. 1986, roč. 30, č. 1, s. 7–18.

SWANSON, Don R., 2008. Literature-Based Discovery? The Very Idea. V: P. BRUZA a M. WEEBER, ed. *Literature-based Discovery*. Berlin: Springer - Verlag Berlin Heidelberg, Information Science and Knowledge Management, 15, s. 3–11. ISBN 978-3-540-68685-9.

SWANSON, Don R. a Neil R. SMALHEISER, 1997. An interactive system for finding complementary literatures: a stimulus to scientific discovery. *Artificial Intelligence* [online]. 1997, roč. 91, č. 2, s. 183–203. Dostupné na: doi:[https://doi.org/10.1016/S0004-3702\(97\)00008-8](https://doi.org/10.1016/S0004-3702(97)00008-8)

SWANSON, D.R., 1988. Migraine and magnesium: eleven neglected connections. *Perspectives in Biology and Medicine* [online]. 1988, roč. 31, č. 4, s. 526–557. Dostupné na: doi:10.1353/pbm.1988.0009

THOMAS, J., E. TOMB, E. THOMAS a G. FAURE, 1994. Migraine treatment by oral magnesium intake and correction of the irritation of buccofacial and cervical muscles as a side effect of mandibular imbalance. *Magnesium Research*. 1994, roč. 7, č. 2, s. 123–127. ISSN 0953-1424.