

Sémantické vyhľadávanie pri rešerši v patentových dokumentoch

Semantic Search in Patent Retrieval

Elena JŮNOVÁ¹

ABSTRAKT

Sémantické vyhľadávanie pri rešerši v patentových dokumentoch je rešeršná technika automatizovaného vyhľadávania údajov, kde cieľom vyhľadávacieho dopytu nie je len nájsť presnú zhodu so zadanými kľúčovými slovami vo vyhľadávaní, ale okrem toho aj určiť zámer používateľa, ktorý vyhľadávanie uskutočnil, a zároveň prideliť kontextový význam vybraným pojmom a termínom, čo v značnej miere rozširuje množinu vyhľadaných odpovedí na zadaný rešeršný dopyt pri zvolených kľúčových slovách. Softvérové algoritmy sémantického vyhľadávania takto do rešeršných dopytov pridávajú určitú úroveň porozumenia, ale zároveň tieto algoritmy implikujú modely učenia (založené na miere odchodov, miere konverzie a pod.), využívajú minulé údaje a vzory pokusov a omylov na zlepšenie používateľského komfortu. V konečnom dôsledku tak sémantické vyhľadávanie zvyšuje spokojnosť používateľov a napomáha lepšiemu priradzovaniu série relevantných odpovedí k zadaným kľúčovým slovám či vybraným dlhším textovým blokom.

ABSTRACT

Semantic patent information retrieval is a search technique of automated data information retrieval, where the goal of the search query is not only to find an exact match with the entered keywords in the patent retrieval, but also to determine the intention of the user, who made the patent retrieval and at the same time to assign contextual meaning to the selected terms, which expands the set of searched answers to the entered search query for the selected keywords. In this way, semantic search software algorithms add a certain level of understanding to search queries. At the same time, these algorithms imply learning models (based on bounce rates, conversion rates, etc.), use past data and models and patterns of trials and errors to improve user experience. Ultimately, semantic patent information retrieval increases user satisfaction and helps to better match a series of relevant answers to entered keywords or selected longer text blocks.

Kľúčové slová

rešeršné vyhľadávanie, sémantické vyhľadávanie, patentové dokumenty, vyhľadávanie podľa kľúčových slov, rešerš v univerzitnom prostredí

Key words

patent retrieval, semantic patent information retrieval, semantic search, patents documents, key words search, patent information retrieval in the university

Úvod

Patentové dokumenty reprezentujú smerovanie ľudskej aktivity vo všetkých sférach, najmä však ekonomickej, technologickej a sociálnej činnosti. V dôsledku zvyšovania dostupnosti technických dokumentov v digitálnom formáte narastá potreba ich efektívneho vyhľadávania. Patentové dokumenty predstavujú bohatý zdroj technických,

právnych a obchodných informácií prezentovaných vo všeobecne štandardizovanom formáte, ktoré často nie sú prezentované v iných dostupných zdrojoch.

Patentové informácie sú dôležitým zdrojom informácií pre výskumníkov a vedcov, podnikateľov, pre organizácie aj obchodné podniky, univerzity a v neposlednom rade patentových odborníkov. Význam patentových dokumentov

¹ Mgr. Elena Jůnová, PhD., Centrum pre transfer technológií, Univerzitný vedecký park, Žilinská univerzita v Žiline, Univerzitná 8215/1, 010 26 Žilina, elena.junova@uniza.sk.

tkvie predovšetkým vo funkcii ochrany, vytvárajú stav techniky a zverejnením informujú verejnosť o pokroku v danej oblasti. Patentová ochrana umožňuje držiteľovi patentu vylúčiť, obmedziť využitie vynálezu tretími stranami alebo ho podmieniť kúpou licencie, a to v teritoriálnom a časovom ohraničení. Zverejnenie patentového dokumentu poskytuje verejnosti prístup k informáciám o nových technológiách. Informácie pomáhajú stimulovať inovácie a prispievajú k hospodárskemu rastu vo svete. Z uvedeného je zrejme, že význam patentových rešerš je veľmi veľký a ich dôležitosť a hodnota v súčasnom svete stále narastá. S neustálym nárastom počtu patentových prihlášok každý rok je potreba efektívnej, systematickej správy vyhľadávania takého obrovského množstva údajov nevyhnutnosťou s mimoriadnym významom.

Článok prezentuje základné princípy sémantického vyhľadávania v patentových dokumentoch, čo môže byť prospešné pre uvedenie do problematiky pre širšiu verejnosť. Zároveň ponúka priblíženie koncepcie sémantického vyhľadávania a poukazuje na rozvoj, možnosti, ale aj limity sémantického vyhľadávania v patentových dokumentoch. Pochopenie systému fungovania a významu spracovania rešerš pri výskume v univerzitnom prostredí prináša prínos pre vedecko-výskumných pracovníkov. Sémantická rešerš pri vyhľadávaní v patentových dokumentoch môže (nielen) pre túto skupinu predstavovať zjednodušenie a sprístupnenie rešeršnej činnosti, ktorá by inak musela byť vynaložená pri práci s patentovými dokumentmi, ktoré si vyžadujú znalosť a orientáciu vo výskumnej problematike, štruktúre patentových dokumentov, v klasifikácii patentového triedenia a v mnohých ďalších okruhoch.

1 Rešerše v patentových dokumentoch

Patentové dokumenty spravidla zverejňujú technologické informácie opisujúce vynálezy v súlade s požiadavkami platného patentového práva a uvedením patentovateľného predmetu, ktorý musí spĺňať podmienku novosti, vynálezcovskej činnosti a priemyselnej využiteľnosti, pričom je súčasťou patentovej prihlášky aj odkaz na súčasný stav techniky. Patentové dokumenty sú takto zdrojom informácií nielen o tom, čo je nové (vynález), ale aj o tom, čo už je známe a v mnohých prípadoch poskytujú zhrnutie histórie technologického vývoja v oblasti, ktorej sa týkajú. Súčasťou zverejneného patentového dokumentu býva publikovanie rešeršnej správy, ktorá odkazuje na súbor nájdených dokumentov, ktoré sa dotýkajú predmetného vynálezu a posudzuje relevantnosť novosti a zásahu do patentových nárokov.

Informácie obsiahnuté v patentovom dokumente sú globálne a sú sprístupnené pre všetkých jednotlivcov alebo organizácie na celom svete, čo umožňuje komukoľvek učiť sa a budovať na tomto poznaní. Patenty a z nich vyplývajúca ochrana stimuluje výskum a je kľúčovou požiadavkou

na získavanie kapitálu. Zásadne ovplyvňujú celkový hospodársky rast. Na účely posúdenia vynálezu či technológie prinášajú patentové dokumenty spravidla tie najnovšie informácie (WIPO, 2021a). Ako uvádza WIPO, dôvodom je, že uchádzači sa ponáhľajú; zvyčajne žiadateľovi, ktorý ako prvý podal žiadosť o udelenie patentu na vynález, bude udelená priorita voči ďalším žiadateľom o podobný vynález a ostatné žiadosti budú zamietnuté. Len s patentom má jeho vynálezca, resp. držiteľ práv k dispozícii maximálne zákonné prostriedky na spochybnenie použitia jeho vynálezu inými proti jeho vôli. V neposlednom rade môže vynálezca s patentom spravidla stanoviť vyššiu predajnú cenu alebo licenčné poplatky za využívanie alebo za predaj vynálezu oproti nepatentovanej technológii, ktorá nie je teritoriálne ani časovo chránená.

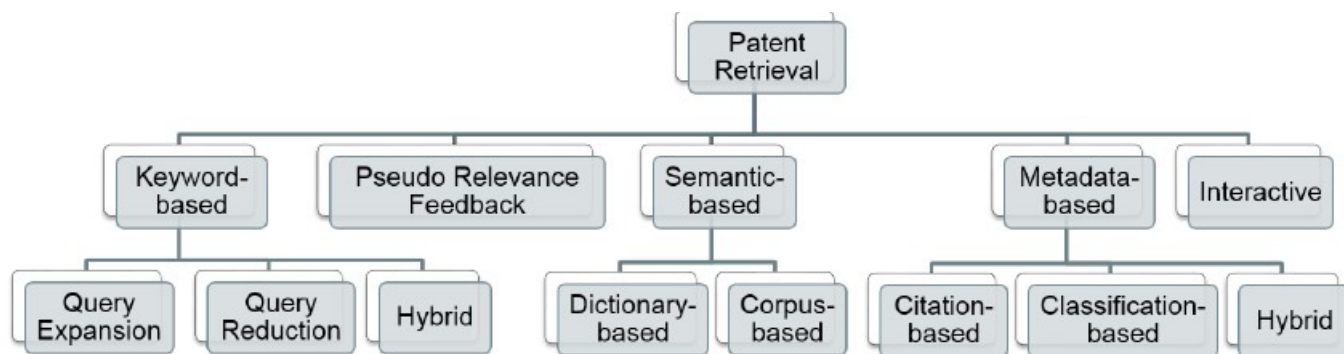
Patentové dokumenty slúžia širšiemu portfóliu používateľov, vďaka rešerši v patentových dokumentoch môžu používateľom pomôcť v rozsiahlom spektre (WIPO, 2021b, Shalaby a Zadrozny, 2018):

- vyhnúť sa duplikácii úsilia v oblasti výskumu a vývoja,
- určiť patentovateľnosť vynálezov,
- vyhnúť sa porušovaniu patentov iných vynálezcov,
- odhadnúť hodnotu vlastných patentov alebo hodnotu patentov iných vynálezcov,
- využiť technológie z patentových prihlášok, ktoré nikdy neboli udelené, neplatné v niektorých krajinách alebo z patentov, ktoré už nie sú platné,
- získať informácie o inovatívnych aktivitách a budúcom smerovaní podnikania od konkurencie,
- zlepšiť plánovanie obchodných rozhodnutí, akými sú licencovanie, partnerstvá v zdieľaní technológií, ale aj fúzie a akvizície,
- prieskum technológie s cieľom zachytiť nové a trendové technológie v konkrétnej doméne a následne ich použiť na vytváranie nových inovatívnych služieb,
- analýza použitia technológie či vynálezu v konkrétnej krajine s cieľom posúdiť rozsah patentovej ochrany špecifickej technológie a následne podľa toho nasmerovať činnosť výskumu a vývoja,
- konkurenčná analýza a benchmarking s cieľom identifikovať silné a slabé stránky vlastného patentového portfólia spoločnosti v porovnaní s inými kľúčovými hráčmi pracujúcimi na príbuzných technológiách.

1.1 Stratégia vyhľadávania v patentovej rešerši

Pri patentovej rešerši je nutné si určiť cieľ prieskumu a vyhľadávania. Podľa účelu a našej potreby potom zameriame naše rešeršné vyhľadávanie. Iný typ rešerše vykonávame, ak vyhľadáваме konkrétny patent, iný, ak nás zaujíma portfólio patentov od konkrétneho vynálezcu či pôvodcu práv, iný, ak zisťujeme teritoriálne obmedzenie, kde daný vynález platí, iný, ak posudzujeme patentovateľnosť vynálezu, s ktorým sami prichádzame na trh, a pod.

Najčastejšie vykonávanou rešeršou je rešerš na stav techniky. Do stavu techniky zahrňame všetky informácie, ktoré



Obr. 1 Metódy patentového vyhľadávania (Shalaby, Zadrozny, 2018)

boli dostupné verejnosti kdekoľvek na svete v akejkoľvek forme k dátumu vykonania rešerše. Pre prihlasovateľa, vývojára či obchodníka bude teda stavom techniky akýkoľvek dôkaz, že jeho myšlienka, nápad alebo produkt je už známy.

Na rešeršné hľadanie v patentovej literatúre a na analýzu patentovej rešeršnej požiadavky sa v odbornej terminológii používa pojem **Patent Retrieval** (PR) alebo tiež **Patent Search**. Patent Retrieval je pilierom takmer všetkých úloh patentovej analýzy, pričom jej podskupina je **Information Retrieval**, ktorá sa zaoberá vývojom techniky a metód, ktoré účinne a efektívne vyhľadávajú príslušný patentový dokument po zadaní dopytu vyhľadávacej požiadavky.

Ako uvádzajú Shalaby, Zadrozny (2018) PR metódy je možné kategorizovať v závislosti od toho, ktoré údaje sa použijú z vyhľadávacích dopytov a z vyhľadávacieho korpusu na získanie príslušných dokumentov. **Metódy založené na kľúčových slovách** používajú iba výrazy z vyhľadávacích dopytov a hľadajú presné zhody v cieľovom korpusu. **Pseudorelevantné metódy spätnej väzby** využívajú termíny z najlepšie umiestnených výsledkov počiatočného dopytu na zlepšenie série relevantných získaných výsledkov. Na základe sémantickej metódy sa analytické nástroje pokúšajú prekonať problém s nesúladom slovníka medzi zadanými vyhľadávanými termínmi v patentových dokumentoch a súvisiacou slovnou zásobou na základe ich významu. **Metódy založené na metadátach** využívajú jazykovo nezávislé netextové metadáta klasifikačného charakteru a bibliografické informácie s cieľom zlepšiť možnosť opätovného vybratia patentu. **Interaktívne metódy** sa napokon zameriavajú na lepšiu organizáciu a prezentáciu výsledkov vyhľadávania pre používateľov. Pri tejto metóde sú používatelia zapojení v interaktívnom procese hľadania, skúmania a očakávania, s cieľom získať čo najviac relevantných výsledkov.

Činnosti súvisiace s prácou s patentovými dokumentmi vyžadujú znalosť a orientáciu v problematike, štruktúre patentových dokumentov, v klasifikácii patentového triedenia a mnohých ďalších okruhoch. Úroveň odborných znalostí sa ešte zvyšuje pri samotnej rešeršnej činnosti, ktorá by mala byť vykonávaná integrovane, sofistikovane,

s použitím inteligentnej analytiky, ktorá poskytuje kognitívnu a interaktívnu pomoc užívateľom. Nakoľko sú patentové dokumenty (Shalaby a Zadrozny, 2018) viacstránkové, multimodálne, viacjazyčné, pološtruktúrované a bohaté na metadáta – tak tieto charakteristiky ich robia vhodnými na vyhľadávanie pomocou inteligentných softvérov a nástrojov.

Cieľom rešeršného vyhľadávania je získať relevantné patentové dokumenty pre dané vyhľadávanie (dopyt, resp. rešeršnú otázku). Táto žiadosť môže mať rôzne formy, ako napríklad: kombinácia kľúčových slov, s použitím Booleovských operátorov a patentového klasifikačného triedenia. Pri použití inteligentných softvérov a nástrojov je možné ako vyhľadávací dopyt zadať poznámku alebo kompletný textový dokument (napr. patentovú prihlášku).

Sémantické hľadanie, ktoré ponúkajú niektoré komerčné softvérové analytické nástroje, ponúka textové vyhľadávanie vo voľnom formáte, s možným použitím automatizovaného strojového textového prekladu. Umožňuje výber konceptu, ktorý funguje zadaním kľúčového slova a spustením vyhľadávania. Potom je možné podrobne rozbaľiť zobrazený strom konceptov a vybrať koncepty relevantné pre hľadaný dopyt. Proces výberu konceptu ďalej vedie k zobrazeniu zoznamu relevantných skupín zoradených na základe ich skóre relevantnosti. Vyhľadávanie podobnosti možno ďalej vykonať v zozname vybraných relevantných skupín pomocou symbolov medzinárodnej patentovej klasifikácie a kooperatívnej patentovej klasifikácie, ako aj konceptov, pomocou ktorých sa podobné patenty nájdu. Zobrazené výsledky je možné vizualizovať formou rôznych grafov a podporných funkcií, pričom výsledky je možné aj zdieľať s inými spolupracovníkmi a pod.

2 Sémantické vyhľadávanie pri rešerši v patentových dokumentoch

Sémantické vyhľadávanie pri rešerši v patentových dokumentoch sa dostáva do popredia laickej verejnosti, ale aj ako nápomocný prostriedok pre odbornú verejnosť v posledných rokoch. Algoritmus sémantického vyhľadávania

je trénovaný, aby v databáze patentových dokumentov, ktoré sú zaznamenané v štandardizovanom informačnom formáte, vedel spracovať zadaný dopyt (zadaný či už vo forme viacerých kľúčových slov, ale aj kompaktnej časti textu či celej patentovej prihlášky) a v priebehu niekoľkých sekúnd vytvoril pre používateľa súbor relevantných výsledkov vyhľadávania. Ak je vstupným dopytom celý patentový dokument, pri analýze a spracovaní bude algoritmus ignorovať prechodné frázy a neurčité pojmy a tiež identifikuje odvodené termíny, ktoré obsahujú predpony alebo prípony. Algoritmus sémantického vyhľadávania je trénovaný na „porozumenie“ základnému významu slov a vybuduje vzťah medzi slovami a prepojenia medzi nimi. Výsledok potom porovná s patentovými dokumentmi uloženými v databáze (ktorých môžu byť rádovo stovky miliónov v závislosti od používaného nástroja a databázy, s ktorou pracuje) a zobrazí pre používateľa najpodobnejšie a najrelevantnejšie výsledky.

Sémantické vyhľadávanie je technika vyhľadávania údajov, kde cieľom vyhľadávacieho dopytu nie je len nájsť presnú zhodu so zadanými kľúčovými slovami vo vyhľadávaní, ale okrem toho aj určiť zámer používateľa, ktorý vyhľadávanie urobil, a zároveň prideliť vybraným pojmom kontextový význam, čo v značnej miere rozšíri množinu vyhľadaných odpovedí na dopyt pri zadaných kľúčových slovách. Sémantické vyhľadávanie takto do dopytov pridáva úroveň porozumenia, ale tieto algoritmy majú zároveň aj modely učenia (ktoré sú založené na miere odchodov, miere konverzie a pod.), využíva minulé údaje a vzory pokusov a omylov na zlepšenie používateľského komfortu. V konečnom dôsledku tak sémantické vyhľadávanie zvyšuje spokojnosť používateľov a napomáha lepšie priradiť k zadaným kľúčovým slovám relevantné odpovede a stránky. Koncept sémantického vyhľadávania predstavila spoločnosť Google v roku 2010. Odvtedy sa ukázalo ako veľmi účinný nástroj, ktorý napomáha nájsť používateľom viac relevantných výsledkov k ich dopytu, ako by vyhľadávacie nástroj našiel len s použitím vyhľadávania na základe kľúčových slov.

Základný rozdiel medzi vyhľadávaním podľa kľúčových slov a sémantickým vyhľadávaním je v tom, že sémantické vyhľadávanie prináša skôr výsledky založené na koncepte použitých kľúčových slov než na presnej zhode s kľúčovými slovami. Sémantické vyhľadávanie ďalej nepočíta iba s opakovaním kľúčových slov a meraním blízkosti pojmov, ako to robí vyhľadávanie kľúčových slov, ale skôr s využitím naprogramovaného algoritmu na predpovedanie a pochopenie kontextového významu fráz. Nesmierny význam môže mať práve vyhľadávanie na sémantickom princípe pri odhalovaní tzv. skrytých informácií. Aj technické pojmy, ktoré sú predmetom vyhľadávania, je možné pomenovať rôznymi pojmami synonymického významu, častokrát práve v patentových dokumentoch nachádzame čo najširšie vymedzenie jednotlivých pojmov, určitú formu zahmlievania (neprežrádzať viac, ako je nevyhnutné),

forma napísania doterajšieho stavu techniky je napísaná zámerne nejasne – v takýchto prípadoch môžu sémantické vyhľadávacie nástroje rýchlo odhaliť skryté informácie. Vo vyhľadávacích databázach tretích strán je k dispozícii niekoľko nástrojov na analýzu podľa doterajšieho stavu techniky, ktoré umožňujú sémantickým vyhľadávateľom zostaviť koncepčne relevantný zoznam doterajšieho stavu techniky. Vyhľadávateľ zadáva nároky alebo konkrétne prvky nároku, abstrakty, frázy alebo číslo patentu predmetu ako dopyt a vyhľadávacie modul vytvorí vzťah medzi týmto a tým, čo existuje v tele doterajšieho stavu techniky, skôr ako doručí zoznam koncepčne vhodných dokumentov podľa doterajšieho stavu techniky. V komerčných aj nekomerčných patentových vyhľadávacích databázach je k dispozícii niekoľko nástrojov na analýzu podľa doterajšieho stavu techniky, ktoré umožňujú sémantickým vyhľadávateľom zostaviť koncepčne relevantný zoznam doterajšieho stavu techniky. Vyhľadávateľ zadáva patentové nároky alebo konkrétne časti patentového nároku, abstraktu, frázy alebo číslo patentu ako dopyt. Vyhľadávacie modul vytvorí vzťah medzi týmto a tým, čo je súčasťou doterajšieho stavu techniky a následne zostaví zoznam koncepčne vhodných dokumentov doterajšieho stavu techniky (Iam-media, 2020).

Súčasná komerčná riešenia sa vyvinuli z tých, ktoré boli k dispozícii na začiatku tohto desaťročia a ktoré sa pokúšali poskytovať takmer automatické riešenia s minimálnym zásahom človeka s použitím latentných sémantických algoritmov (LSA), na rozdiel od tých, ktoré sú k dispozícii teraz, ktoré akceptujú a dokonca povzbudzujú užívateľa k špecifikovaniu jeho požiadavky a tak získania spätnej väzby pre lepšie výsledky. Extrakcia konceptov, kde je možné vyjadriť termín jedným slovom (napr. dron) alebo ako terminologickú sekvenciu (napr. *bezpilotné lietadlo*) a zaradenie vhodnej slovnej zásoby, ktorá identifikuje, že tieto výrazy zodpovedajú rovnakému významu, by mala poskytnúť lepšie výsledky a presnosť, ktoré sú teraz bežnejšie v sémantických vyhľadávaníach druhej generácie.

2.1 Princípy sémantického vyhľadávania v patentových dokumentoch

Rešeršné vyhľadávanie založené na kľúčových slovách môže byť tvorené rešeršnou otázkou, ktorá je zložená len z niekoľkých pojmov, ale rovnako môže byť rešeršná otázka v sémantickom vyhľadávaní pri patentovej rešerši tvorená aj celým patentovým dokumentom alebo jeho časťou, čo môže predstavovať aj tisíce výrazov. Medzi základné metódy používané pri sémantickom rešeršnom vyhľadávaní patrí **slovníková metóda** a **korpusová metóda**. Ich aplikácia býva interdisciplinárna, prepájajúca obe metódy pre lepšie výsledky vyhľadávania.

Pri oboch metódach sa používa metóda preformulovania rešeršnej otázky (Query reformulation – QRE) prostredníctvom zníženia alebo rozšírenia termínov Q na účely zvýšenia relevantnosti nájdených dokumentov (Shalaby a Zadrozny, 2018):

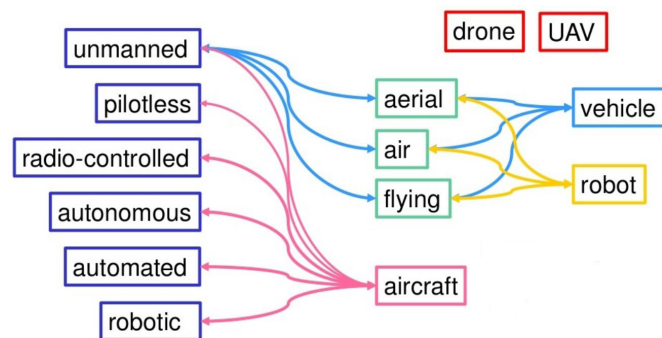
- **Query Reduction (QR):** Redukcia rešeršnej otázky v rešeršnom vyhľadávaní predstavuje účinný spôsob najmä pri takom vyhľadávaní, kde rešeršná otázka tvorí celý patentový dokument alebo jeho časť. Vtedy je prehľadávanie všetkých použitých termínov neefektívne. Tam, kde termínom zo špecifických častí alebo z časti patentového dokumentu je pridelená väčšia váha než ostatným termínom, je metóda založená na pozícii pojmu v štruktúre dokumentu. Ďalším príkladom redukcie rešeršnej otázky sú metódy založené na patentovom triedení IPC² a CPC³, ktoré navyše používajú termíny z definícií IPC a CPC ako lexikón alebo zoznam stop-slov.

- **Query expansion (QE):** Rozšírenie rešeršnej otázky na základe syntaktických pravidiel jazyka. Expanzné metódy fungujú tak, že expandujú termíny s výrazmi podobného významu, ako sú synonymá alebo hyponymá.

- **Spojenie Query expansion a Query reduction:** V tejto hybridnej metóde sú irelevantné výrazy odstránené z rešeršnej otázky a relevantnejšie výrazy sú pripojené k rešeršnej otázke.

Sémantické rešeršné vyhľadávanie založené na slovníkovej metóde: Sémantické rešeršné vyhľadávanie založené len na kľúčových slovách sa ukázalo byť nedostatočne účinné kvôli nesúladu slovníka medzi podmienkami rešeršnej otázky a relevantným obsahom patentov. Shalaby a Zadrozny (2018): v zbierke CLEF-IP 2009⁴ až 12 % príslušných dokumentov neobsahovalo bežne používané pojmy vo vyhľadávaných témach. To zvyšuje potrebu nových prístupov na preklenutie medzier v slovníkoch medzi používanými termínmi. Najčastejšie sú pri slovníkovej metóde sémantického vyhľadávania pri rešerši využívané slovníky WordNet⁵, TechNet⁶, WIPO Pearl⁷ a iné druhy technických tezaurových slovníkov. Slovníková metóda má svoje limity. Na základe týchto poznatkov a zavedených metód je možné čiastočne automaticky identifikovať synonymá/ekvivalenty, súčasne sa vyskytujúce výrazy a určiť blízkosť vzťahov na rozšírenie dopytu. Nápomocná je hlbšia analýza protokolov rešeršnej otázky zvažujúca iné metadáta, ako sú relevantné prístupy početností, napríklad na základe polohy, frekvencie, časti, kde sa termín nachádza a pod.

Sémantické rešeršné vyhľadávanie založené na korpusovej metóde: Aj korpusová metóda je založená na preformulovaní rešeršnej otázky QRE na báze sémantiky. Pri týchto metódach sa analyzujú textové korpusy, aby sa extrahovali sémanticky súvisiace pojmy s výrazmi dopytu, ktoré je možné použiť na rozšírenie rešeršnej otázky (Shalaby a Zadrozny, 2018).



Obr. 2 Kolekcia príbuzných termínov s pojmom „dron“ (Kirch-Verfuß, 2016)

Vo viacerých odborných publikáciách na tému sémantickej rešerše v patentových dokumentoch nachádzame ako príklad uvedený rešeršný dopyt termínom „dron“ a „UAV (bezpilotný letecký prostriedok)“. Na obrázku 2 vidíme kolekciu termínov spájajúcich sa s vyhľadávaným termínom „dron“ a „UAV“ a väzbami a prepojeniami medzi jednotlivými termínmi, pričom kolekcia môže mať ešte širšiu koncepciu a množstvo vyhľadaných výrazov, ako vidíme na obrázku. Na tomto príklade si môžeme vysvetliť **základný princíp sémantického vyhľadávania v patentových dokumentoch** (Kirch-Verfuß, 2016). Preto uvádzame príklad v anglickom jazyku, nakoľko väčšina sémantického vyhľadávania je implementovaná na anglický jazyk, resp. dochádza iba k prekladu z iných jazykov do angličtiny, ak je takáto možnosť sémantického vyhľadávania v ponuke.

2 IPC – Medzinárodné patentové triedenie (z angl. International Patent Classification)

3 CPC – Kooperatívne patentové triedenie (z angl. Cooperative Patent Classification)

4 CLEF-IP 2009 je elektronická databáza patentových dokumentov, ktoré boli extrahované z databáz UPSTO, EPO a WIPO. Patentové dokumenty boli prevedené do XML formátu a schémy Document type definition (DPD). Táto databáza je rozdelená na dve časti, a to korpusovú časť a tematickú časť.

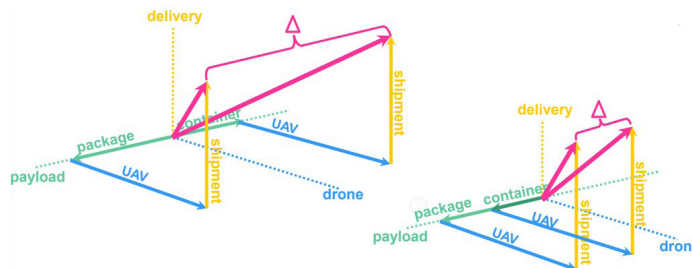
5 WordNet – je rozsiahla lexikálna databáza anglického jazyka. Podstatné mená, slovesá, prídavné mená a príslovky sú zoskupené do súborov kognitívnych synonym (synsetov), z ktorých každé vyjadruje odlišný pojem. Synsety sú prepojené pomocou konceptuálno-sémantických a lexikálnych vzťahov. Je to užitočný nástroj na výpočtovú lingvistiku a spracovanie prirodzeného jazyka. WordNet sa povrchne podobá na tezaurus, pretože zoskupuje slová podľa ich významov. Existuje však niekoľko dôležitých rozdielov. Po prvé, WordNet neprepája len slovné formy – reťazce písmen – ale konkrétne zmysly slov. Výsledkom je, že slová, ktoré sa nachádzajú v sieti navzájom v tesnej blízkosti, sú sémanticky disambigantné. Za druhé, WordNet označuje sémantické vzťahy medzi slovami, zatiaľ čo zoskupenia slov v tezaure nesledujú žiadny iný explicitný vzorec než významovú podobnosť (WordNet, 2021).

6 TechNet – sémantická elektronická databáza orientovaná na technologické poznatky. Vychádza z amerických patentových dokumentov USPTO (Sahrica a kol. 2018).

7 WIPO Pearl – je viacjazyčný terminologický portál WIPO. WIPO Pearl poskytuje prístup k vedeckým a technickým pojmom odvodeným z patentových dokumentov. Pomáha podporovať presné a konzistentné používanie výrazov v rôznych jazykoch a uľahčuje vyhľadávanie a zdieľanie vedeckých a technických znalostí (WIPO, 2021c).

- V prvom rade sa hľadajú **morfologické variácie pojmu** – skloňovanie, plurál a pod. („dron“, „drones“, „dron´s“ a pod.)
- **Rozlišovanie slovného zmyslu**, v angl. word-sense disambiguation (WSD) – zameranie na identifikáciu toho, aký význam pojmu sa používa vo vete:
 („Bagpipe drones (US6476301)“)
 („Trapper and retainer for bee swarms and separator for drones and queens (US1896925)“)
 („Autonomus and automatic landing system for drones (US2009055038)“)
- **Hľadanie synonym** – hľadanie pojmov, ktoré znejú inak, ale majú rovnaký význam ako daný pojem:
 („unmanned aerial vehicle“)
 („pilotless aircraft“)
 („flying robot“)
- **Hľadanie homoným** – hľadanie pojmov, ktoré znejú rovnako, ale majú odlišný význam ako daný pojem:
 („male bee“)
 („pipe of bagpipe“)
 („unchanging monotone“)
 („aircraft without a pilot“)
- **Hľadanie hypernym** – zovšeobecnenie alebo generalizácia samotného pojmu:
 („Means of transportation:
 ..Vehicle ..Aerial vehicle ... unmanned aerial vehicle
 ..Craft – vehicle designed for navigation in air, water, outer space ...aircraftpilotless aircraftradio-controlled aircraft“)
- **Konceptuálne prepájanie**, v angl. concept matching – extrakcia konceptu z vyhľadávaného textu a následné

spájanie týchto extrahovaných konceptov z textu z dokumentu do skupín.



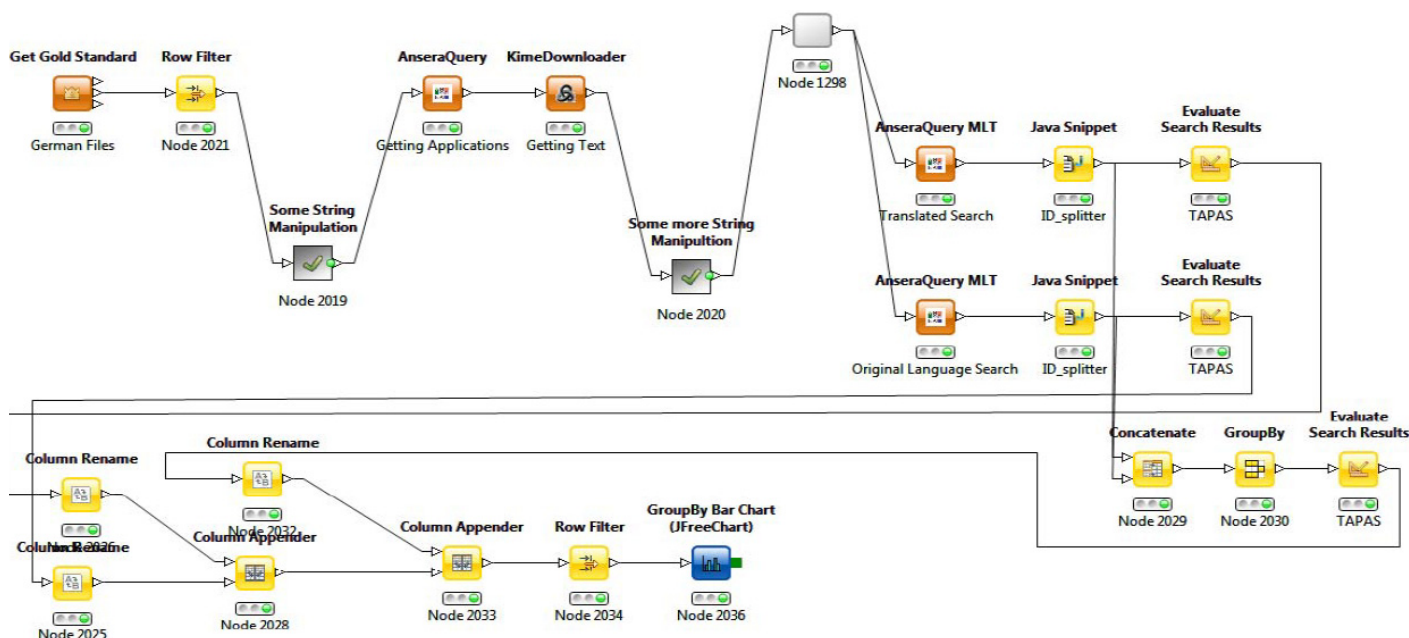
Obr. 3 Spájanie extrahovaných konceptov z textu dokumentu do skupín (Kirch-Verfuß, 2018)

V kontexte nešpecifikovaného lietajúceho vozidla (v angl. unmanned aerial vehicle – UAV) na termíny „package“ a „container“ môžeme nazerať ako na podobné (Kirch-Verfuß, 2018).

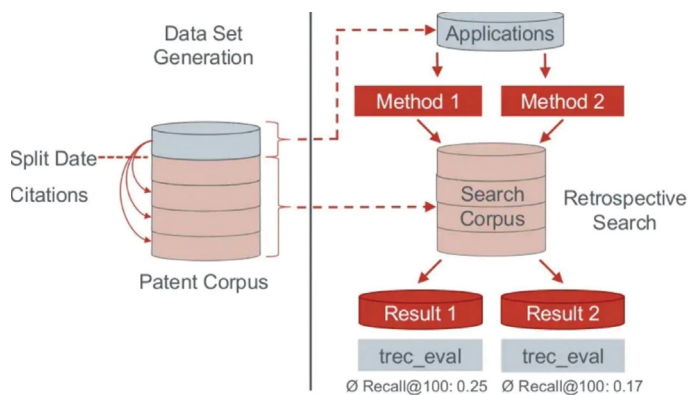
• **Vedomostné prepájanie**, v angl. knowledge matching – zapájanie inteligentných vedomostných nástrojov, napr. ontologických, a následné hodnotenie pridaných dokumentov.

Na konferencii usporiadanej Európskym patentovým úradom (EPO) v roku 2017 predstavil koncept sémantického vyhľadávania pri rešerši v patentových dokumentoch Alexander Klenner-Bajaja. Samotný koncept rozčlenil na procesy:

• **Benchmarkingový porovnávací systém vyhľadávania**
 Počas procesu sémantického vyhľadávania v patentových dokumentoch dochádza k tzv. **benchmarkingu**. V tomto prípade je benchmarkingový proces porovnávania zameraný na vyhľadávanie a prostredie, v ktorom sa nachádzajú



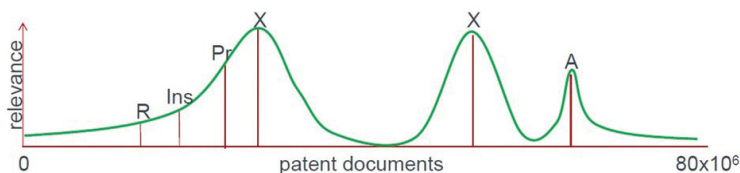
Obr. 4 Holistický prototyp benchmarkingového systému vyhľadávania (Klenner-Bajaja, 2017)



Obr. 5 Benchmarkingový systém vyhľadávania s použitím štatistických metód (Klenner-Bajaja, 2017)

termíny. Ide o pomerne zložitý systém využívajúci viaceré nástroje. My ako funkčný ilustračný príklad uvádzame schému vyhľadávacieho benchmarkingu, aký bol používaný v Európskom patentovom inštitúte (EPO) v roku 2017.

Ako vidíme na obrázkoch 4 a 5 v benchmarkingovom prostredí sa počas porovnávacieho procesu využívajú štatistické metódy a zložité algoritmy. Výsledkom benchmarkingového vyhľadávania v patentovej literatúre je rešeršná správa, ktorej predchádzalo vyhľadávanie vo viac ako 40 miliónoch patentových rodín. V rešeršnej správe sú spomínané relevantné dokumenty označované kategóriou **X** alebo **A** citácie/odkazy. Iba pár citácií obsahuje kategóriu X alebo A. V priemere sú to 2 citácie v X-kategórii a 3 citácie v A-kategórii. Vyhľadávanie však umožňuje aj zobrazenie ďalších relevantných dokumentov, informácií a odkazov v kategóriách **R** – vrátené počas vyhľadávania, **Ins** – odložené na podrobnú kontrolu, **Pr** – tlačené, **Cs** – uložené na citáciu (Klenner-Bajaja, 2017), pozri obrázok 6.



Obr. 6 Benchmarkingové vyhľadávanie a zobrazenie výsledkov z rešeršnej správy v kategóriách (Klenner-Bajaja, 2017)

• **Klasifikácia patentového triedenia na presnejšie vyhľadanie**

Zatriedenie do medzinárodného klasifikačného triedenia IPC, CPC prispieva k zobrazeniu relevantnejších výsledkov.

B	PERFORMING OPERATIONS; TRANSPORTING
B64	AIRCRAFT; AVIATION; COSMONAUTICS
B64C	AEROPLANES; HELICOPTERS
B64C39/00	Aircraft not otherwise provided for
B64C39/02	. characterised by special use
B64C39/024	. . of the remote controlled vehicle type, i.e. RPV
B64C39/028	. . Micro-sized aircraft

Obr. 7 Zatriedenie hľadaného termínu „dron“ do CPC (Kirch-Verfuß, 2016)

• **Kľúčové slová a generovanie rešeršných otázok v (automatických) scenároch vyhľadávania a automatické rozšírenie rešeršných otázok**

Generovanie rešeršných otázok v automatických scenároch a automatizované rozšírenie vyhľadávacieho dopytu môžeme znázorniť na obrázku 8. Rozšírenie sa deje na základe korpusovej metódy, v tomto prípade Word2Vec⁸ s použitím prekrývajúcich sa pojmov a zároveň rozšírenie vyhľadávateľných termínov na základe automatického návrhu podloženého znalosťou prepojenia väzieb dopytovateľných termínov. Pre každý termín je pritom vypočítaná váha a nadväznosť na pôvodný termín.

▪ Word2Vec Expansion Terms in Overlap

New Overlap	Original Keyword	Word2Vec Enrichment
outputs	signal	OR signals OR circuit OR output OR clock OR outputs
apparatus	controlling	OR control OR controlled OR apparatus
screen	display	OR screen OR displays OR displaying OR displayed OR lcd
increasing	increases	OR increase OR decreases OR increased OR increasing OR decrease
serially	decreasing	OR increasing OR decrease OR increase OR decreases OR decreased
	parallel	OR serially

▪ Re-Suggestion of known Query Terms

Term	Original Keyword	Word2Vec Enrichment
transmission	transmitted	OR transmits OR received OR transmit OR transmission OR transmitting
	reception	OR transmission OR transmitted OR transmits
	transmitting	OR receiving OR transmitted OR transmission OR sending OR transmit

Obr. 8 Automatické rozšírenie vyhľadávacej rešeršnej otázky (Klenner-Bajaja, 2017)

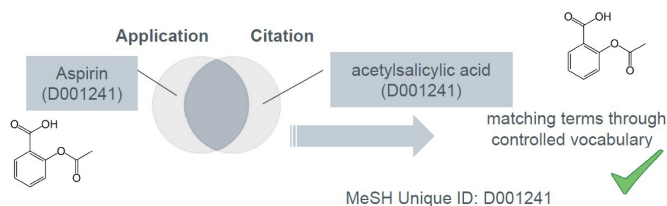
• **Sémantika založená na terminológii**

Sémantika založená na terminológii predstavuje v rámci projektu APL⁹ anotáciu celej zbierky doterajšieho stavu

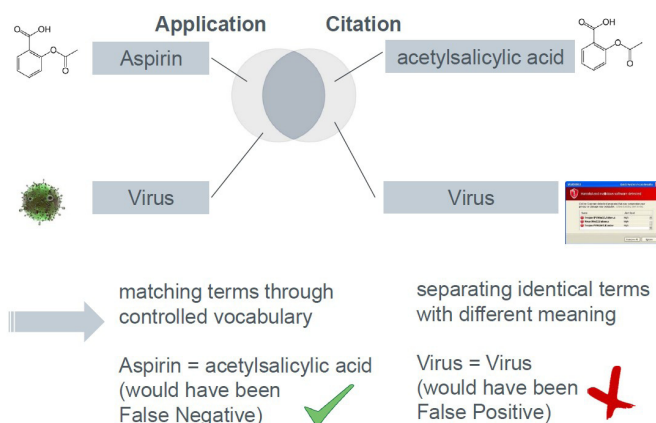
8 Word2Vec – je technika na spracovanie prirodzeného jazyka publikovaná v roku 2013. Algoritmus Word2Vec používa model neurónovej siete na učenie asociácií slov z veľkého korpusu textu. Po vyškolení môže taký model detegovať synonymické slová alebo navrhnúť dodatočné slová k čiastkovému dopytu. Word2Vec pridružuje jednotlivým pojmom konkrétne vektorové čísla. Vektory sú vyberané starostlivo tak, aby jednoduchá matematická funkcia (kosínusová podobnosť medzi vektormi) naznačovala úroveň sémantickej podobnosti medzi slovami reprezentovanými týmito vektormi.

9 APL – je programovací jazyk používajúci matematicky odvodené symboly pre mnohé zo svojich operácií. APL je interaktívny jazyk orientovaný a integrovaný na vývojové prostredie, usposobený pre väčšinu počítačových platforiem.

techniky v normalizovanej podobe zameranú na chemické entity, fyzikálne jednotky, citácie, riadené terminológie tak, aby bolo možné sémantické vyhľadávanie týchto entít, pozri obrázok 9. Na obrázku 10 je zachytený homonymický rozpor v identicky znejúcich výrazoch s rozdielnym významom, pričom bol algoritmus schopný tento rozpor vyhodnotiť ako falošne pozitívny výsledok.



Obr. 9 Sémanticky zjednotená terminológia (Klenner-Bajaja, 2017)



Obr. 10 Sémanticky zjednotená terminológia s identifikáciou falošne pozitívnych výsledkov (Klenner-Bajaja, 2017)

• Dôvera v automatizované výsledky vyhľadávania

Jedným z prostriedkov na validáciu a zhodnotenie výsledkov automatizovaného vyhľadávania je predikčný model citlivosti na základe distribučných štatistík z tepelných „heat“ máp či 10-násobná krížová validácia (Y-scrambling).

2.2 Rozvoj, možnosti a limity sémantického vyhľadávania v patentových dokumentoch

Limitujúcim, ale aj rozvojovým faktorom sémantického vyhľadávania v patentových dokumentoch je spracovanie prirodzeného jazyka (NLP). Na rozvoji NLP sa stále pracuje, je podmienené zberom obrovského množstva dát v textovej podobe, ktoré učia stroje zákonitostiam jazyka, jeho väzbám a prepojeniam. Počítačové systémy prirodzeného jazyka musia prevádzať vhodným spôsobom informácie

z počítačových databáz do prirodzeného jazyka používaného ľuďmi, čo je náročný proces. Morfológický model reči musí byť schopný generovať gramaticky správne vety, porozumieť významu slov, kontextu, a prípadne musí byť schopný ďalšieho prekladu do iného jazyka. Značné problémy spôsobuje samotná flektívnosť, teda mnohotvarosť jazykov (pričom aj slovenský jazyk patrí medzi flektívne jazyky), ktorá sa do umelej inteligencie ťažko implementuje, ale zároveň predstavuje príležitosť na skvalitnenie a hlbšiu analýzu. Zatiaľ je najlepšie spracovaným jazykom v podobe NLP anglický jazyk, v podobe ktorého sa nachádza veľké percento patentovej literatúry. Spracovávanie a vylepšovanie ďalších jazykov rieši množstvo IT tímov, organizácií, spoločností po celom svete s čoraz lepšími výsledkami vďaka inteligentným algoritmom. Pracuje sa na automatickom strojom preklade jazykov a tiež na prevode znakových súborov latinkových a nelatinkových. V nelatinkových znakových súboroch je písané množstvo patentovej literatúry, pričom práve výskum a vývoj v Číne, Japonsku a Kórei predstavuje značný potenciál a širokú základňu patentovej literatúry.

Zároveň sa pri sémantickom vyhľadávaní musíme popasovať s nejednotnosťou terminológie. Moja „škatuľka“ je váš „kartón“, resp. môj „dron“ je váš „bezpilotný letecký prostriedok“ a rovnaká terminologická variabilita platí pre pojmy prevažnej väčšiny. Príčiny tejto nekonzistentnosti sú spôsobené rôznymi pohľadmi rozličných vynálezcov, žiadateľov, patentových zástupcov či regionálnymi rozdielmi v terminológii a tiež rôznymi jazykmi. Navyše niekedy žiadatelia a patentoví zástupcovia úmyselne používajú neurčitý jazyk na skrytie zámeru vynálezu alebo niekedy na to, aby vynález vyzeral sofistikovanejšie, ako je („príklad – polygonálna nádoba na báze recyklovateľných vlákien“). Pomôcť môžu klasifikačné symboly patentov IPC alebo CPC, ale niekedy aj jednoduché objekty môžu mať veľa symbolov. Pred použitím týchto symbolov ako obmedzenia vo vyhľadávacom dopyte je potrebné vytvoriť dlhý zoznam príslušných symbolov a kódov (Lloyd, 2019).

V texte v prirodzenom jazyku sa dajú identifikovať nasledujúce javy, ktoré odrážajú kontextovú zapojenosť jazykových prvkov a sú dôležité na správne spracovanie termínov a pojmov reprezentujúcich obsah textu (Paralič a kol., 2010), (Tóth, 2008):

• Zámená, kontextové odkazy, anaforické referencie, odvolávky

V texte sa pomerne často vyskytujú konštrukcie obsahujúce odkazy na objekty z predchádzajúceho alebo nasledujúceho kontextu. Určenie takýchto kontextových črt je pri spracovaní textu pomerne ťažké zachytiť. Napr. vo vete „Kľúčové slová a ich použitie v rešeršnej otázke.“ zámeno ich odkazuje na kľúčové slová. Priradenie zámena k príslušnému termínu vyžaduje vysokú úroveň syntaktickej a sémantickej analýzy textu.

- **Synonymia**

Slová s odlišnou formou a rovnakým alebo podobným významom (napr. *kniha, publikácia, dokument*) by sa mali reprezentovať tým istým termínom. Na identifikáciu podobnosti významov je potrebná slovotvorná, sémantická a pragmatická analýza.

- **Homonymia, polysémia**

Slová s náhodne totožnou formou a s rôznym významom (napr. *jazyk = reč vs. jazyk = v ústach, vs. jazyk = na topánke* a pod.) by sa mali podľa kontextu rozlíšiť a reprezentovať vzájomne rôznymi termínmi. Na rozlíšenie polysémických slov treba uskutočniť syntaktickú a sémantickú analýzu textu.

- **Frázy, viacslovné ustálené pomenovania**

Slovné spojenia so špecifickým významom, napr. *manažment znalostí, expertný systém* a pod., majú charakter samostatných termínov. Ich celkový význam je iný ako významy slov, z ktorých sa tieto termíny skladajú. Bolo by teda vhodné, aby sa ustálené viacslovné pomenovania reprezentovali jedným viacslovným termínom.

- **Porozumenie významu textu**

Ak by analyzovaný text obsahoval tvrdenie „*Nejde o dron so snímajúcim zariadením*“, tak by reprezentácia obsahu tohto textu nemala zahŕňať termíny snímajúce zariadenie. Takýto stupeň porozumenia významu textu, ak je dosiahnuteľný, by si vyžadoval úplnú sémantickú a pragmatickú analýzu, s využitím znalostí mimojazykových skutočností.

- **Nejednoznačnosti pri spracovaní textu vo fázach tokenizácie a lematizácie** (najmä pri flektívnych jazykoch) – napr. slovotvorný základ termínu mier (možné lemy: *mier, miera, mieriť*)

Na jednoznačné rozhodnutie, ktorá z možných lem je správna, je potrebné zistiť syntaktickú štruktúru danej vety, z nej odvodiť prípustné gramatické kategórie vetných členov a na ich základe určiť správnu lemu.

- **Určenie neplnovýznamových slov, tzv. stop-slová** (angl. stop-words)

Plnia len syntaktické funkcie. Tieto slová možno vylúčiť z ďalšej analýzy.

- **Určenie kľúčového slova, ktoré nie je obsiahnuté v spracovávanom texte**, ale je možné ho identifikovať pomocou istých deduktívnych a inferenčných nástrojov, patrí tiež medzi zložitú úlohu kontextovej analýzy.

Uvedené okruhy môžu byť príčinou nepresností pri spracovaní prirodzeného jazyka, a tak následne dôsledkom nepresností pri samotnom rešeršnom vyhľadávaní založenom na sémantickom prístupe vyhľadávacími strojmi,

zároveň ako limitujúci faktor môžeme vnímať zlé zatriedenie do patentového klasifikačného triedenia (IPC, CPC) spolu s dopytovanými termínmi či textom, resp. absenciu toho zatriedenia, kde konečný sumár výsledkov môže byť touto skutočnosťou výrazne ovplyvnený. Preto vo všeobecnosti platí, že i keď používateľ môže počas tohto procesu stratiť transparentnosť viazanú k výsledkom vyhľadávania (slepé dôverovanie algoritmom bez poznania logiky, ktorá za nimi stojí) a kontroly (v podobe autority či znalca nad konštruovaním rešeršnej otázky), získava na vnútornej väzbe extrahovanej medzi koncepciou a dokumentmi.

Aj v odborných kruhoch je považované sémantické vyhľadávanie pri rešerši v patentových dokumentoch za žiadanú výpomoc. Vďaka prelínaniu znalostí a skúseností rešeršných špecialistov reflektujúce na vystavení rešeršnej otázky, môže nástroj sémantického vyhľadávania reprezentovať kvalitatívne vyššiu efektívnosť, a tak v konečnej podobe súbor nájdenej relevantnej patentovej literatúry môže byť ešte širší. Rešeršní profesionáli pri formulácii rešeršnej otázky kooperujú s reťazcom strategicky vybraných kľúčových slov a logických operátorov a začlenenia do patentového klasifikačného triedenia. Ostatným, ktorí nemajú osvojenú túto pokročilú zručnosť, môže napomôcť sémantický vyhľadávací nástroj prehľadávať skupinu údajov založených skôr na význame ako na samotných kľúčových slovách. Sémantické vyhľadávanie humanizuje nepríjemný proces štruktúrovania rešeršných otázok podobný robotu. Vyhľadávač odvodzuje kontext, spája významy a identifikuje dokumenty obsahujúce príslušné pojmy. IP (2021) poukazuje na to, že tento druh vyhľadávania umožňuje nájsť spojenia, o ktorých používatelia sémantických vyhľadávacích služieb možno ani netušili, že vôbec existujú. Tieto vzťahy sú objavené prostredníctvom spracovania prirodzeného jazyka a algoritmov strojového učenia. Umelá inteligencia umožňuje vyhľadávaču identifikovať význam v rámci rešeršnej otázky a obsahu databázy, ktorý má byť vyhladaný. Táto technológia predstavuje obrovský pokrok od vyhľadávacích nástrojov založených na kľúčových slovách, kde obmedzené vstupy viedli k obmedzenému nachádzaniu výsledkov.

2.3 Prínosy sémantického vyhľadávania pri rešerši v patentových dokumentoch v univerzitnom prostredí

V univerzitnom prostredí prebieha výskum a vývoj v širokom spektre odvetví. Ale oproti bežnému trhovému prostrediu, kde sa spoločnosti orientujú na trhové presadenie a implementáciu svojich výrobkov na trh, sa špecifická univerzitného výskumu ešte prehľbujú. Spoločnosti využívajú rešerše v patentových dokumentoch s cieľom zachytiť nové a trendové technológie vo sfére svojho pôsobenia a následne použiť tieto technológie na vytváranie nových inovatívnych služieb. Rešeršná analýza býva vykonávaná aj so zacielením na použitie konkrétnej technológie či

vynálezu vo zvolenej krajine s cieľom posúdiť rozsah patentovej ochrany a následne podľa toho nasmerovanie činnosti výskumu a vývoja, resp. umiestnenie výrobku na trh danej krajiny. Rešeršná analýza môže byť použitá ako konkurenčná analýza s cieľom identifikovať silné a slabé stránky patentového portfólia vlastnej spoločnosti v porovnaní s inými kľúčovými hráčmi venujúcimi sa príbuzným technológiám. Rešeršná analýza vedie k zlepšeniu plánovania obchodných rozhodnutí, akými sú licencovanie, partnerstvá v zdieľaní technológií, ale aj fúzie a akvizície. Takéto rešeršné analýzy si spoločnosti buď dávajú vypracovať externými rešeršnými profesionálmi, alebo zamestnávajú rešeršného špecialistu, ktorý vypracováva rešeršné analýzy (či už tematické – na stav techniky alebo na novosť, alebo bibliografické, alebo na patentovú čistotu).

V prípade univerzitných výskumníkov je rešeršná analýza tiež mimoriadne prospešná a žiadaná. Zárukou kvality výskumu v univerzitnom prostredí je novosť a uplatniteľnosť výsledkov výskumu do praxe. Z uvedeného dôvodu Čorejová, Jaroš, Hornická (2021) zdôrazňujú kladenie dôrazu na správne nakladanie s výsledkami výskumu v univerzitnom prostredí, ich efektívnu ochranu a komercializáciu. Poukazujú aj na to, že to vyžaduje interné nastavenie súvisiacich procesov na univerzitách, s dôrazom na definovanie podmienok, práv a povinností účastníkov týchto procesov, ako aj rozsahu a obsahu súvisiacich dokumentov. Rešeršnú činnosť považujú za nosnú súčasť komplexného procesu ochrany a komercializácie výsledkov výskumu v univerzitnom prostredí. Upriamujú pozornosť najmä na význam spracovania rešerší na stav techniky za účelom identifikácie patentových dokumentov, ktoré by mohli byť námietkou proti novosti alebo vynálezcovskej činnosti predmetu rešerše, a teda jeho úspešnej ochrany a uplatnenia do praxe. Rovnako však poukazujú aj na to, že rešeršná činnosť slúži aj pre potreby zamestnávateľa ako podklad na rozhodovanie v procese ochrany a komercializácie výsledkov výskumu, najmä uplatnenia si práva na riešenia vytvorené v zamestnaneckom režime, posúdenie vhodnosti zvolenej formy a územného rozsahu ochrany predmetu rešerše, pri odhade komerčného potenciálu a pod., kde univerzita vystupuje ako subjekt, ktorý daný výskum financuje a očakáva z neho prínos. V snahe o efektívne nakladanie s výsledkami výskumu a vedeckého bádania boli zriadené na univerzitách na Slovensku a v zahraničí za týmto účelom centrá pre transfer technológií. V rámci svojej činnosti tieto centrá poskytujú podporu a poradenstvo výskumníkom, zamestnancom univerzity či študentom univerzít, v procese ochrany a komercializácie nimi vytvorených výstupov charakteru duševného vlastníctva.

V univerzitnom prostredí v prvotnej fáze výskumu, pred začatím výskumu alebo v jeho počiatočných štádiách, nemusia výskumníci využiť externé zdroje na vykonanie rešerše na stav techniky v danej výskumnej sfére, ale ju zabezpečia vlastnými internými zdrojmi. Najčastejšie môže ísť o nadväzujúci výskum v oblasti, ktorej sa venujú

buď samotní výskumníci, doktorandi, študenti, alebo je naň orientovaná univerzitná katedra, alebo môže ísť o podnet z externého prostredia, resp. spoluprácu v určitej oblasti s komerčným prostredím a jeho špecializovanými požiadavkami a pod. Vykonávanie rešerše v patentových dokumentoch interným spôsobom v podobe vyhľadávania v patentových dokumentoch samotnými výskumníkmi prináša určité úskalia.

Činnosti súvisiace s prácou s patentovými dokumentmi vyžadujú znalosť a schopnosť orientovať sa v problematike, štruktúre patentových dokumentov, v klasifikácii patentového triedenia a v mnohých ďalších okruhoch. Práve tu môže vystúpiť do popredia koncept sémantického vyhľadávania v patentových dokumentoch, kde napriek jeho finančnej náročnosti prevažujú prínosy pre jednoduchosť používania a rýchle získanie relevantných výsledkov. Algoritmus sémantického vyhľadávania je trénovaný, aby v databáze patentových dokumentov, ktoré sú zaznamenané v štandardizovanom informačnom formáte, vedel spracovať zadaný dopyt (zadaný či už vo forme viacerých kľúčových slov, ale aj kompaktné časti textu či celej patentovej prihlášky) a v priebehu niekoľkých sekúnd vytvoril pre používateľa súbor relevantných výsledkov vyhľadávania. Ak je vstupným dopytom celý patentový dokument, pri analýze a spracovaní bude algoritmus ignorovať prechodné frázy a neurčité pojmy a tiež identifikuje odvodené termíny, ktoré obsahujú predpony alebo prípony. Algoritmus sémantického vyhľadávania je trénovaný na „porozumenie“ základnému významu slov a vybuduje vzťah medzi slovami a prepojenie medzi nimi. Výsledok potom porovná s patentovými dokumentmi uloženými v databáze (ktorých môžu byť rádo vo stovky miliónov v závislosti od používaného nástroja a databázy, s ktorou pracuje) a zobrazí pre používateľa najpodobnejšie a najrelevantnejšie výsledky. Sémantické vyhľadávanie je technika vyhľadávania údajov, kde cieľom vyhľadávacieho dopytu nie je len nájsť presnú zhodu so zadanými kľúčovými slovami vo vyhľadávaní, ale okrem toho aj určiť zámer používateľa, ktorý vyhľadávanie urobil, a zároveň prideliť vybraným pojmom kontextový význam, čo v značnej miere rozšíri množinu vyhľadaných odpovedí na dopyt pri zadaných kľúčových slovách. Práve takéto sémantické vyhľadávanie v patentových dokumentoch môže napomôcť výskumníkom určiť smerovanie ich záujmu a orientácie na novú oblasť výskumu, znížiť redundantnosť a nehospodárnosť vývoja už vyvinutých technológií či viesť k rozhodnutiu implementácie vyvinutej technológie alebo vyvinutého vynálezu do trhovej oblasti, či k samotnému elementárnemu rozhodnutiu, či sa daná technológia či vynález oplatí chrániť formou patentu, resp. úžitkového vzoru a v akom teritoriálnom rozsahu.

3 Záver

V článku boli prezentované základné princípy sémantického vyhľadávania v patentových dokumentoch, priblížil koncepciu sémantického vyhľadávania s poukazaním na

ďalší možný rozvoj, ale aj limitujúce faktory sémantického vyhľadávania v patentových dokumentoch. Článok zároveň priblížil špecifické postavenie potreby rešeršnej činnosti v univerzitnom prostredí a priblížil možné alternatívy rešeršnej činnosti výskumníkov v ich výskume a bádani.

Nástroje sémantického vyhľadávania sú založené na algoritmoch umelej inteligencie a pomáhajú objavovať relevantné údaje, zlepšovať prístup k neustále sa zvyšujúcim objemom informácií. Tým, že sémantická technológia prístupuje k automatickému chápaniu významov, prekonáva hranice ostatných technológií. Expert AI Team (2018) označuje za najlepší spôsob porozumenia tomu, ako môžu „inteligentné“ nástroje ponúkať takú výhodu, je porozumenie tomu, ako sa líšia od typických prístupov k vyhľadávaniu. Tradičné metódy vyhľadávania založené na kľúčových slovách alebo štatistických algoritmoch hľadajú na základe relevancie a frekvencie kľúčových slov. Jednoducho nerozumejú významu slov v kontexte. Čisté technológie strojového učenia, ktoré môžu byť v niektorých prípadoch nápomocné, si vyžadujú veľa času a obrovské množstvo úsilia na to, aby boli vyškolené (a priebežne sa školili, aby zostali účinné aj v priebehu času). Strojové učenie sa dá použiť iba na základné úlohy, keď nie je potrebná inteligencia, pretože nemá vstavané znalosti. Sémantické vyhľadávacie nástroje založené na umelej inteligencii namiesto toho využívajú pokročilé kognitívne funkcie na porozumenie ľudskému jazyku a extrahovanie zmysluplných informácií z akejkoľvek znalostnej základne, webovej stránky alebo patentových dokumentov. Sémantické vyhľadávanie pri rešerši v patentových dokumentoch je druh patentového vyhľadávania, ktoré prináša výsledky založené na konceptoch kľúčových slov, nie na presných zhodách kľúčových slov z rešeršnej otázky. Namiesto zhody kľúčových slov sa zameriava na význam a podobnosť sémantického obsahu, asociácie s kľúčovými slovami – zretazovanie, združovanie predstáv, pri ktorom jedna predstava za určitých podmienok vyvolá druhú na základe podobnosti alebo protichodnosti, resp. predstava založená na základe predchádzajúcej skúsenosti. Sémantické vyhľadávanie je obzvlášť užitočné na širšie, koncepcnejšie vyhľadávanie a tiež na prieskumy v patentovej oblasti. V prípade, že patentovú rešerš nevypracováva profesionál, je zväčša oveľa rýchlejšia a relevantnejšia ako vyhľadávanie na základe kľúčových slov. Sémantické vyhľadávanie v patentových dokumentoch je tiež nápomocné pri všeobecných patentových rešeršiach, pretože výber termínov a pojmov v podobe kľúčových slov aj pri zhodnej riešenej výskumnej oblasti býva v rôznych patentových dokumentov rôznorodý. Pomocou správneho sémantického vyhľadávacieho nástroja je možné tieto „skryté informácie“ rýchlo nájsť. Sémantické vyhľadávanie je účinné aj v rade ďalších patentových rešerší vrátane zhodnotenia patentovateľnosti vynálezu či jeho zneplatnenia, či pri samotnej rešerši na doterajší stav techniky.

Literatúra

- AMBERCITE. 2021. [online]. [cit. 2021-08-03]. Dostupné na: <https://www.ambercite.com/>
- ANASOFT. 2021. Sémantická analýza – AIIA – Medzipodniková interoperabilita [online]. [cit. 2021-06-11]. Dostupné na: <https://www.anasoft.com/sk/home/inovacie/veda-a-vyskum/inovacie-aktualne-projekty/semanticka-analyza-aiia-%E2%80%93-medzipodnikova-interoperabilita>
- ARAS, H., TURKER, R., GEISS, D., MILBRADT, M., SACK, H. 2018. Get Your Hands Dirty: Evaluating Word2Vec Models for Patent Data [online]. In: CEUR-WS.org. VOL – 2198. [cit. 2021-05-10]. Dostupné na: <https://fizweb-p.fiz-karlsruhe.de/sites/default/files/FIZ/Dokumente/Forschung/ISE/Publications/Aras-Semantics-2018.pdf>
- BUKOVČANOVÁ, A. 2021. Natural language processing alebo spracovanie prirodzeného jazyka je čoraz podstatnejšie [online]. In: Transdata. [cit. 2021-05-15]. Dostupné na: <https://www.translata.sk/blog/natural-language-processing-alebo-spracovanie-prirodzeneho-jazyka-je-coraz-podstatnejsie>
- CIARAMELLA, M., CIARAMELLA, A. 2019. What kind of semantics and why to use it in patents: part 1 – Semantic search [online]. In: Hal. VOL-02092622 [cit. 2021-07-19]. Dostupné na: <https://hal.archives-ouvertes.fr/hal-02092622/>
- DONG, H., KHADEER HUSSAIN, F., CHANG, E. 2008. A Survey in Semantic Search Technologies [online]. In: Conference: Digital Ecosystems and Technologies, 2008. DEST 2008. 2nd IEEE International Conference. DOI:10.1109/DEST.2008.4635202. [cit. 2021-06-16]. Dostupné na: https://www.researchgate.net/publication/224331268_A_survey_in_semantic_search_technologies
- ČOŘEJOVÁ, A., JAROŠ, J., HORNICKÁ, K. 2021. Význam internej rešerše na stav techniky v univerzitnom prostredí [online]. In: TRANSFER TECHNOLÓGIÍ bulletin. [cit. 2022-09-15]. Dostupné na: <https://ttb.sk/clanky/vyznam-internej-reserse-na-stav-techniky-v-univerzitem-prostredii/>
- EPO. 2019. Linked open EP data - User guide [online]. In: Electronic Publication and Dissemination [cit. 2021-07-22]. Dostupné na: [http://documents.epo.org/projects/babylon/eponet.nsf/0/A42D42FBF4DC69C6C1258216005B5AAF/\\$File/linked_open_ep_data_user_guide_en.pdf](http://documents.epo.org/projects/babylon/eponet.nsf/0/A42D42FBF4DC69C6C1258216005B5AAF/$File/linked_open_ep_data_user_guide_en.pdf)
- EXPERT AI TEAM. 2018. 3 awesome benefits of AI-based semantic search tools for your business [cit. 2021-08-26]. Dostupné na: <https://www.expert.ai/blog/3-awesome-benefits-ai-based-semantic-search-tools-business/>
- HELMERS, L., HORN, F., BIEGLER, F., OPPERMAN, T., MULLER, K., R. 2019. Automating the search for a patent's prior art with a full text similarity search [online]. In: ArXiv. VOL-1901.03136 [cit. 2021-07-15]. Dostupné na: <https://arxiv.org/pdf/1901.03136.pdf>

- IAM-MEDIA. 2020. How semantic search/automated patent analysis tools can save researchers time [online]. In: Effectual Knowledge Services Pvt Ltd - IP technology [cit. 2021-05-10]. Dostupné na: <https://www.iam-media.com/how-semantic-search-automated-patent-analysis-tools-can-save-researchers-time>
- INTEPAT. 2019. Steps to Conduct Google Patents Search [online]. [cit. 2021-07-18]. Dostupné na: <https://www.intepat.com/blog/patent/google-patents-search/>
- IP. 2021. Increase Intelligence around IP with Semantic Search: Best Practices for IP Professionals Series [online]. In: eBook Innovation Q Plus [cit. 2021-07-18]. Dostupné na: https://ip.com/wp-content/uploads/2020/09/IQ_Semantic_Search_eBook_P2.pdf
- KIRCH-VERFUß, G. 2016. Searching for Information – the Classical Way with Key Words and Classification compared with a Semantic Approach [online]. In: WissensWert Tutorial [cit. 2021-05-10]. Dostupné na: <https://www.slideshare.net/Haxel/ivic-2016-tutorial-searching-for-information-the-classical-way-with-key-words-and-classification-compared-with-a-semantic-approach-67573367>
- KIRCH-VERFUß, G. 2018. Semantic search versus searching with terms and classifications – what helps best in patent information searching. [cit. 2021-05-10]. In: EPO Patent Information Conference. [ppt presentation]. WissensWert Seminare – Beratung.
- KLENNER-BAJAJA, A. 2017. Toward semantic search at the European Patent Office [online]. [cit. 2021-05-10]. EPO Patent Information Conference – Data Scientist Requirements Engineering – Solution Design. Brussels. In: SlideShare. Dostupné na: <https://www.slideshare.net/Haxel/iisdv-2017-towards-semantic-search-at-the-european-patent-office>
- KRACKER, M. 2018. Linked Open EP data – a new Product from the EPO [online]. EPO – EU Datathon. In: SlideShare. Nice: 2018. [cit. 2021-05-10]. Dostupné na: <https://www.slideshare.net/Haxel/icsdv-2018-martin-kracker-epo-linked-open-ep-data-a-new-product-from-the-epo>
- LLOYD, M. 2019. Ambercite Searching: How does it compare to semantic searching? – A patent office comparison [online]. In: Amber blog. [cit. 2021-05-19]. Dostupné na: <https://www.ambercite.com/amberblog/2019/06/11/ambercite-patent-office-assessment>
- MINESOFT. 2021. Minesoft add powerful Semantic Search module to PatBase [online]. [cit. 2021-06-26]. Dostupné na: <https://minesoft.com/2016/11/28/minesoft-release-semantic-search-module-patbase/>
- PARALIČ, J., FURDÍK, K., TUTOKY, K., BEDNÁR, P., SARNOVSKÝ, M., BUTKA, P., BABIČ, F. 2010. Dolovanie znalostí z textov [online]. In: Technická univerzita v Košiciach [cit. 2021-08-01]. ISBN 978-80-89284-62-7. Dostupné na: <http://people.tuke.sk/jan.paralic/knihy/DolovanieZnalostizTextov.pdf>
- PATBASE. 2016. A summary of the latest enhancements to PatBase and PatBase Express [online]. In: What's new? Minesoft. [cit. 2021-07-12]. Dostupné na: <https://www.patbase.com/pb201610.pdf>
- PATSEER. 2019. ReleSense™ – An AI-driven NLP engine built for patent and scientific literature [online]. [cit. 2021-07-12]. Dostupné na: <https://patseer.com/2019/04/semantic-patent-search-analysis-relesense/>
- PATSEER. 2021. [online]. [cit. 2021-08-03]. Dostupné na: <https://patseer.com/>
- PATSNAP. 2021. What is AI connected innovation intelligence? [online]. In: Why PatSnap. [cit. 2021-07-12]. Dostupné na: <https://www.patsnap.com/why-patsnap/>
- QUESTEL ORBIT INTELLIGENCE. 2021. Unlock hidden insights in global Patent Database [online]. [cit. 2021-07-12]. Dostupné na: <https://www.questel.com/business-intelligence-software/orbit-intelligence/>
- SERHARD, S., JIANXI, L., WOOD, K., L., 2020. TechNet: Technology Semantic Network Based on Patent Data [online]. In: ScienceDirect. Elsevier. VOL 142, 112995 [cit. 2021-06-18]. Dostupné na: <https://www.sciencedirect.com/science/article/abs/pii/S0957417419307122>
- SHALABY, W., ZADROZNY, W., 2018. Patent Retrieval: A Literature Review [online]. In: ArXiv. VOL 1701.00324. [cit. 2021-08-01]. Dostupné na: <https://arxiv.org/pdf/1701.00324.pdf>
- SULEMAN, E. 2021. Alternatives to Google Patents [online]. In: Trainindex. [cit. 2021-08-04]. Dostupné na: <https://www.trainindex.io/blog/alternatives-to-google-patents-4j4b/>
- TAY, A. 2019. The rise of the „open“ discovery indexes? Lens.org, Semantic Scholar and Scinapse [online]. [cit. 2021-08-01]. Dostupné na: <http://musingsaboutlibraries.blogspot.com/2019/12/the-rise-of-open-discovery-indexes.html>
- TÓTH, J. 2008. Kontextová analýza textu [online]. Autoreferát dizertačnej práce. In: Slovenská technická univerzita v Bratislave. [cit. 2021-06-28]. Dostupné na: <https://www.fe.stuba.sk/buxus/docs/2013/autoreferaty/Toth.pdf>
- YOO, J. Y., MORRIS, J. X., LIFTLAND, E., QI, Y. 2020. Searching for a Search Method: Benchmarking Search Algorithms for Generating NLP Adversarial Examples [online]. In: ArXiv. VOL-2009.06368 [cit. 2021-06-29]. Dostupné na: <https://arxiv.org/pdf/2009.06368.pdf>
- WIPO, 2021a. Patent information and development [online]. [cit. 2021-07-10]. Dostupné na: https://www.wipo.int/edocs/mdocs/sme/en/wipo_wasme_ipr_ge_03/wipo_wasme_ipr_ge_03_5-main1.pdf
- WIPO, 2021b. The Role of Patent Information in Supporting Innovation [online]. [cit. 2021-07-10]. Dostupné na: https://www.wipo.int/edocs/mdocs/sme/en/wipo_smes_rom_09/wipo_smes_rom_09_e_workshop02_1-related1.pdf
- WIPO, 2021c. WIPO Pearl – WIPO's Multilingual Terminology Portal [online]. [cit. 2021-07-12]. Dostupné na: <https://www.wipo.int/reference/en/wipopearl/>
- WORDNET. 2021. What is WordNet? [online]. In: Princeton University. [cit. 2021-06-19]. Dostupné na: <https://wordnet.princeton.edu/>