Patenty na VŠ v ČR

Obě se zajímáme o ekonomii, což nás přivedlo k tomu, abychom se zaměřily na téma inovací – konkrétně na patenty českých vysokých škol. Chtěly jsme zmapovat vysoké školy a finanční přínos jim plynoucí z patentů a v jakých oblastech mají největší motivaci patentovat. Tento projekt propojil naši vášeň pro ekonomickou efektivitu, vědu a analytické myšlení. Ekonomika, inovace a trocha detektivky – to je přesně náš styl!

Získání dat

Úřad duševního vlastnictví poskytuje datové sady ve formátu \*.xml, kdy publikace probíhá formou plného exportu databáze a následných přírůstků. Znělo to skvěle, stáhnout pár souborů a máme hotovo! ALE! Data nelze vybrat na základě časového období, které chceme zkoumat a musíme stáhnout vše. No, jenže každý zip soubor, obsahuje asi 100 složek se 100 \*.xml soubory a k tomu plno obrázkových příloh. Paměť PC nestačí – musíme na to chytře přes Python. Stáhni soubor zip, který extrahuj, ale do složky nahraj jen xml soubory.

Máme celkem 1 737 619 souborů ve formátu \*.xml. Přichází na řadu porada s Chat GPT, který navrhne parsování \*.xml souborů s pomocí knihoven xml.etree.ElementTree. Na řadu přichází i pandas či knihovny csv a os. Parsujeme data s podmínkou, že datum přihlášky je větší 2013 a výsledné vybrané hodnoty ukládá do \*.csv souboru. Data následně čistíme a mažeme přihlášky, které nepodaly veřejné vysoké školy. Vzhledem k tomu, že variant zápisu škol je tolik, je nejrychlejší úprava v Excelu. Následně bylo třeba tabulku normalizovat a zbavit se některých duplicit, pro které byly vytvořeny vazební tabulky. Jeden patent totiž mohlo podat více spolupracujících škol a týkají se více oblastí zájmu. Pro dočištění dat jsme použili Azure Data Studio. Jedna faktovka byla hotová.

Trápení pokračovalo s parsováním dat z \*.pdf souborů. Zkoumáme totiž licenční příjmy univerzit, jejichž hodnotu získáme z výročních zpráv vysokých škol. Původně byla naivní představa, že ChatGPT zvládne po relativně krátkém ladění promptu postahovat si sám ze zadané URL adresy vysoké školy Výroční zprávy za 10 leté období v \*.pdf formátech a sám vyhodí výstup v podobě úhledné Excelovské tabulky. Po konzultaci s lektorkou Terkou se zdál náš úkol pro AI příliš komplexní a tak jsme ručně postahovaly Výroční zprávy z webových stránek vysokých škol. Všechny výroční zprávy obsahují řádek s hodnotou licenčních příjmů v tabulce označené jako A1. Pro extrakci dat, jsme zkoušely použít knihovny PyPDF2, PDFplumber i jiné. Na řadu přišel i REGEx, se kterým nám pomohla mentorka Adéla. Nicméně po 8 hodinovém trápení na prvním Hackatonu se všechna data nepodařilo stáhnout. Vylepšováním skriptu se něco podařilo, ale i tak jsme část byly nuceny naházet ručně.

Po prvním Hackatonu jsme měly již každá dost dat ve svých laptopech a vyvstala potřeba kooperovat a verzovat. Z nabízejících se možností jsme vybraly GitHub, odzkoušený již s domácích úkolů, který se jevil jako v klidu. Po hlubším zkoumání branchí se nám na druhém hackatonu však nepodařilo rozjet jejich system a tak jsme přispívaly do mainu.

Data jsme nahrály do Power BI a abychom si nemusely data ukládat lokálně, nechaly jsme datové sety v Azure Data Studio databázi, což se osvědčilo v momentě, kdy jsme ještě objevily nějakou nesrovnalost v datech a v Azure Data Studio pomocí jazyka SQL upravily data. Následný update v Power BI pak byl jen otázkou chvilky.

Datový model

Datový model, jsme si původně vyrobily dle našich představ o tématu. Po hlubším seznámení s datasetem jsme udělaly patřičné úpravy a níže uvádíme hotový model.

Trocha teorie

Pokud si chcete váš vynález nechat patentovat, musíte se jednat o **technické řešení**. Jinak řečeno, za návod k technickému řešení, které má být dosaženo technickými prostředky. **Patenty** se udělují na vynálezy, které jsou nové, jsou výsledkem vynálezecké činnosti a jsou průmyslově využitelné. Maximální doba ochrany patentu je 20 let.

Pro vynález si můžete zvolit ochranu **užitným vzorem**, který někdy bývá považován za jednodušší variantu patentu. Oproti patentu je ochrana užitným vzorem rychlejší a méně nákladná. Maximální doba platnosti užitného vzoru je 10 let (standardní 4 roky, které si můžete dvakrát prodloužit na celkových 10 let).

Patent udělený Evropským patentovým úřadem tzv. Evropský patent má stejné účinky jako národní patent. Za udržování evropského patentu v platnosti je majitel povinen platit udržovací poplatky stejně jako u národního patentu.

Patenty na školách

Jelikož patent či užitný vzor musí být technickým řešením, je očekávané, že nejvíce přihlášek podávají univerzity ČVUT, VŠB-TUO a VUT. Překvapením je například Akademie múzických umění v Praze s celkovým počtem 12 přihlášek (jedním z nich je například zařízení ke sledování provozu varhan). Celkem 4 univerzity nepodaly žádnou přihlášku a to Akademie výtvarných umění v Praze (AVU), Janáčkova akademie múzických umění (JAMU), Slezská univerzita v Opavě (SU) a Vysoká škola ekonomická v Praze (VŠE).

Školy využívají ochranu vynálezů především pomocí užitného vzoru (62%) než patentů či evropského patentu (celkem 38%), což je pro školy levnější. Celkem bylo podáno 4 339 přihlášek, z nichž jen 233 nebylo uznáno a patent nebyl udělen, což poukazuje pouze na 95% úspěšnost žadatelů.

Protože s patenty souvisí povinnost hradit udržovací poplatky, část patentů je ukončena předčasně. Udržovací poplatky jsou s každým rokem užívání patentu vyšší. V prvním roce činí 1000 CZK, ve 13. roce užívání již 10 000 a ve 20. roce užívání až 24 000 CZK. Za zkoumané desetileté období bylo předčasně zrušeno 620 patentů. Nejvíce jich rušilo České vysoké učení technické V Praze - 13 %, Vysoká škola báňská - Technická univerzita Ostrava - 15%, Vysoké učení technické v Brně 17%. Nejvyšší míru předčasného ukončení patentů vykazuje Akademie múzických umění v Praze 33%, ale i z důvodu že má pohých 12 přihlášek.

Podíl patentů a vzorů ukončených před vypršením jejich platnosti zůstává víceméně stabilní, což naznačuje, že předčasné ukončování není hlavním faktorem celkového poklesu patentové aktivity. Dlouhodobý pokles celkového počtu patentů však může svědčit o snížení efektivity patentového systému nebo o změně přístupu a strategií českých vysokých škol při využívání patentové ochrany.

Majitelé přinášeli inovace především v oblastech fyziky, chemie, zdraví, záchrany života, zábavy, tváření, elektřiny či staveb.

Licenční příjmy univerzit

Nepodařilo se nám získat data za roky 2014 - 2016 z Českého vysokého učení technického , i když jsme oslovily univerzitu přímo. Univerzity nemají povinnost zveřejňovat výroční zprávy o hospodaření starší 5 let na svých webových stránkách a do uzávěrky projektu nám data bohužel neposkytli.

Své prvenství si drží České vysoké učení technické v Brně také ve výši licenčních příjmů v absolutním vyjádření za sledované období. V těsném závěsu se umístila Mendelova univerzita v Brně a Vysoké učení technické v Brně. Můžeme sledovat, že od roku 2016 se univerzitam velmi daří a licenční příjmy mají rostoucí charakter. Celkem za 10 leté období všechny univerzity tj. 26, měly příjmy z licenčních poplatků ve výši 193 mln.

Průměrný roční příjem všech univerzit je 744 485 CZK. Nejvýkonnější, již zmiňovaná univerzita České vysoké učení technické v Praze, průměrně za rok inkasuje více než 3,5 mln CZK, Vysoké učení technické v Brně jen o něco málo a to průměrně ročně 3,1 mln CZK dále Masarykova univerzita 2,6 mln CZK. Jelikož medián příjmů všech univerzit je o mnoho nižší než průměr a to jen 146 500 CZK dovedlo nás ke kategorizaci příjmů. Rozdělili jsme příjmy do tří kategorii přičemž hranice jsme určili přibližně podle již zmíněného mediánu a průměru. Největší variabilitu příjmů pozorujeme v kategorii Hight Income, kde hodnoty příjmů sahají od středních milionových částek až po extrémní hodnoty (outliery) přesahující 5 mln. Do této kategorie spadá Technická univerzita v Liberci, která v roce 2014 dosáhla příjmů z licencí 6,3 mln či Jihočeská univerzita v Českých Budějovicích, která v roce 2023 měla příjmy z licencovaných patentů 5,3 mln.

Držitel patentu se může rozhodnout zda bude svůj patent licencovat či nikoliv. Důvodem pro ponechání si patentu bez licence je získání konkurenční výhody, udržení si monopolu nad výrobou či distribucí produktu nebo bezpečností či etické důvody. Naopak pro přístup licencování za poplatek se rozhodnou subjekty, které chtějí získat příjmy bez nutnosti vyrábět produkt samotný.

Poskytování licencí je jedním z prostředků k získání finančních příjmů nebo jiné formy prospěchu z výsledků vynálezecké činnosti či předmětů průmyslového vlastnictví (patent, užitný vzor atd.).

Vysoký počet patentů nemusí nutně znamenat vyšší licenční příjmy co vyjadřuje slabou závislost mezi počtem licencovaných patentů a příjmy. Mendelova univerzita v Brně a Západočeská univerzita v Plzni dosahují významných příjmů i s relativně nízkým počtem patentů. Nejúspěšnější v komerializaci svých výzkumných činností je již dříve zmiňované České vysoké učení technické v Praze. Problémy s efektivní monetizací vidíme u Vysoké školy báňské - Technické univerzity Ostrava.

V daném období bylo licencováno celkem 104 patentů podaných v letech 2014 - 2023, z toho 23 patří Českému vysokému učení technickému v Praze, 22 - Vysoká škola báňská - Technické univerzity Ostrava, překvapivě 11 - Univerzita Tomáše Bati ve Zlíně a Vysoké učení technické v Brně - 9 z celkově 199 přihlášek. Počet licencovaných patentů v průběhu času klesá, dalo by se tedy říci, že licencované vynálezy nejsou pro školy příliš zajímavé.

Počet patentových přihlášek od roku 2014 výrazně klesá, což může naznačovat snižující se aktivitu českých vysokých škol v oblasti podávání patentů. Tento trend však může být spojen s rostoucím zaměřením na kvalitu a komerční potenciál patentů, což potvrzuje rostoucí výše licenčních příjmů. Univerzity tak mohou investovat více do strategických patentů s vyšším ekonomickým přínosem místo zaměření na pouhý počet přihlášek. Nicméně klesající počet patentových přihlášek naznačuje, že do budoucna nelze očekávat výrazný růst licenčních příjmů. Omezený počet nových patentů by mohl v dlouhodobém horizontu omezit finanční udržitelnost tohoto přístupu.

Kterým směrem by se dalo jít dál?

Je třeba zdůraznit, že české veřejné vysoké školy generují příjmy i prostřednictvím smluvního výzkumu pro podniky či jiné instituce. Tyto výzkumy jsou zpravidla cílené na specifické potřeby zadavatele, pro něž jsou vyvíjeny. Při nahlédnutí do výročních zpráv jsme zjistily, že právě smluvní výzkum je pro vysoké školy finančně atraktivnější díky výrazně vyšším příjmům. Do budoucna by mohlo být zajímavé tuto oblast podrobněji prozkoumat.

Závěr

Za posledních 10 let absolutní jedničkou co do výše příjmů, počtu licencovaných či nelicencovaných patentů si drží České vysoké učení technické v Praze. Další úspěšné instituce jsou Vysoké učení technické v Brně, Vysoká škola báňské - Technická unierzita Ostrava, Mendelova unierzita, Univerzita Tomáše Bati, Technická Univerzita v Liberci. Licenční příjmy univerzit za posledních 10 let rostly, co je výsledek aktivní vědecké činnosti minulosti před rokem 2013. Do budoucna lze však očekávat, že příjmy budou klesat z důvodu klesající tendence licencovaných patentových žádostí. Výrazný propad v celkovém počtu patentových žádostí lze vysvětlit změnou metodiky hodnocení Vědy, výzkumu a inovací. Podle metodiky platné pro roky 2014 - 2018 byla důležitá kvantita podaných přihlášek, což se však postupně mění a v nové metodice již rozhoduje kvalita a využitelnost patentovaných inovací nad kvantitou. Ačkoli pro školy může patentování výsledků výzkumu přinášet jistou prestiž a uznání jako inovátoři v daných oborech, není pravděpodobně toto pro školy hnacím motorem. Je možné, že se více soustředí na aktivity, které přinášejí vyšší ekonomický užitek prostřednictvím smluvních výzkumů na zakázku.

Jak jsme si to rozdělily?

Iva Němečková

Již před nástupem do Digitální Akademie mě lákalo programování, proto jsem si chtěla vyzkoušet hlavně práci s Pythonem. S pomocí chatu GPT jsem napsala několik kódů pro stahování dat z Úřadu průmyslového vlastnictví. Stahovala jsem stovky souborů v zip formátu, extrahovala je a ukládala. Další kódy byly zaměřené na webscraping databáze Úřadu nebo stahování stavu přihlášek pomocí poskytnutého API. Další využití Pythonu bylo pro práci s xml soubory, ze kterých jsem parsovala vybraná data do csv souborů, abychom je mohly lehce nahrát do databáze. Proces zahrnoval opakované úpravy skriptu, kdy jsme s chatem ladili jeho funkčnost. Chat GPT byl super pomocník, protože mi byl schopen vysvětlit, jak webscraping funguje a jaká data musím získat z vývojářského prostředí webovky. Se základní znalostí procesu pak bylo jednodušší navrhnutý kód kontrolovat a případně žádat úpravu. Po scrapingu jsem se věnovala normalizaci dat a odstranění duplicit, k čemuž jsem následně použila Excel, Azure Data Studio a Power Query. Na základě takto připravených dat jsem navrhla datový model a vytvořila ho v Lucidchartu. V Power BI jsem pak vytvořila vazby mezi tabulkami a navrhla několik přehledných vizualizací. Celý projekt jsme zakončily společným psaním blogového článku s Beou, kde jsme sdílely klíčové poznatky.

Beata Řeháková

Generativní umělá inteligence mě nesmírně fascinuje a snažím se ji využívat na denní bázi. Původně jsem si naivně myslela, že extrakce jednoho čísla z \*.pdf souboru bude hračka. Opak byl však pravdou. ChatGPT mi vracel neuspokojivé výsledky, a tak jsem nakonec výroční zprávy stáhla ručně. S pomocí mentora Honzy jsem následně parsovala \*.pdf soubory pomocí Pythonu, přičemž některé části bylo potřeba zpracovat manuálně. Založila jsem server s databází v Azure Data Studio, kam jsem nahrávala tabulky příjmů, žadatelů a další data, která webscrapovala Ivča. Následně jsem pracovala na čištění, nahrávání a mazání tabulek pomocí SQL v Azure Data Studio nebo Excelu. Vytvořila jsem několik vizualizací v Power BI a podařilo se mi také zpracovat Box Plot v Jupyter Notebooku. Měla jsem velkou chuť naučit se práci s GitHubem, proto jsem si vytvořila účet a snažila se přispívat alespoň do hlavní větve (main). Na blogovém článku jsme spolupracovaly s Ivčou.

Poděkování

Rády bychom poděkovaly celému týmu Czechitas, který se podílel na organizaci podzimního běhu Digitální Akademie Data v Brně. Naše díky patří mentorovi Honzovi Kotenovi, Adél Bierské za rady s Regexem, Martinu Kovářovi za podporu na obou hackathonech. Velké poděkování patří také našim manželům Vláďovi a Martinovi za jejich podporu během celého kurzu. A v neposlední řadě děkujeme všem účastnicím za skvělý spirit, sdílení zkušeností a vzájemnou podporu – bylo to fajne baby!

Zdroje a nástroje

* GitHub celý projekt: <https://github.com/bejka2012/Patenty_BI>
* Úřad průmyslového vlastnictí: <https://upv.gov.cz/>
* Webové stránky vysokých škol v ČR: <https://msmt.gov.cz/vzdelavani/vysoke-skolstvi/prehled-vysokych-skol-v-cr-3>
* Azure Data Studio (SQL)
* VS Code (Python)
* Power BI (Power Query)
* GitHub
* Lucidchart
* ChatGPT
* Excel, Google produkty