Patenty na VŠ v ČR

Fascinace výkonem ekonomiky nás přivedla k tomu, abychom se zaměřily na téma inovací – konkrétně na patenty českých vysokých škol. Chtěly jsme zjistit, jak efektivně dokáží školy monetizovat své nápady prostřednictvím licencování, a zároveň jsme zkoumaly, v jakých oblastech mají největší motivaci patentovat. Tento projekt propojil naši vášeň pro ekonomickou efektivitu, vědu a analytické myšlení. Práce s daty nám umožnila nejen pochopit, jak funguje proces patentování, ale také nahlédnout do oblasti, kde se věda setkává s praxí. 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 poradí, že \*.xml soubory se parsují 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. Pro zvědavé možno shlédnout zde(https://github.com/bejka2012/Patenty\_BI).

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 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 své technické řešení je možné si zvolit ochranu užitným vzorem. Užitný vzor někdy bývá považován za jednodušší variantu vynálezu. Oproti patentové ochraně je ochrana užitným vzorem rychlejší a méně nákladná. Maximální doba platnosti užitného vzoru, při placení poplatků za prodloužení platnosti zápisu, je oproti platnosti patentu poloviční, tj. 10 let (4 roky a můžete si dvakrát prodloužit ochranu na celkem 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, není překvapením, 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 Akademie výtvarných umění v Praze (AVU), Janáčkova akademie múzických umění (JAMU), Slezská univerzita v Opavě (SU), 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 5% 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 se jednalo o 620 patentů, což představuje České vysoké učení technické V Praze - 13 %, Vysoká škola báňská - Technická univerzita Ostrava - 15%, Vysoké učení technické v Brně 17%. Nejvyšší míru vykazuje Akademie múzických umění v Praze 33%, ale i z důvodu že má pohých 12 přihlášek.

? graf

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.). K realizaci této komercializace průmyslového práva a duševního vlastnictví slouží licenční smlouva. Licencí se tedy poskytuje nabyvateli licence právo k výkonu průmyslových práv, tj. například vynález chráněný patentem vyrábět a obchodovat s vyrobenými výrobky. Toto právo, například patent, se tedy licencí neprodává, ale zůstává i nadále v majetku poskytovatele licence. Ve výroční zprávě nelze zjistit, z kterých konkrétních patentů plyne daný příjem, nicméně nějaký přehled máme z databáze patentů.

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. Problemy 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 patentované vynálezy nejsou prozatím pro firmy příliš zajímavé.

Musíme ještě podotknout, že české veřejné vysoké školy získávají příjmy také smluvním výzkumem s podniky či jinými institucemi. Tyto výzkumy bývají laděné spíše na míru konkrétnímu subjektu, s kterým spolupracují a pro koho vyvíjejí. Při zběžném pohledu do výročních zpráv jsme zjistily, že smluvní výzkum je pro vysoké školy zajímavější s ohledem na daleko vyšší příjmy z nich plynoucí.

Náš závěr je, že ačkoli pro školy může patentování výsledků výzkumu přinášet jistou prestiž a uznání jako inovátoři v daných oborech, není pravděpodobně toto pro školy hnacím motorem. Více se soustředí na aktivity, které přinášejí vyšší ekonomický užitek. Pokud chcete studovat technickou školu, může být pro studenta zajímavým zjištěním, jestli se škola věnuje inovacím a jestli se za dobu studia dostane k technologiím budoucnosti.

Iva Němečková

Již před nástupem do Digitální Akademie jsem se zabývala automatizací a zjednodušováním práce pomocí Pythonu, což mě přirozeně vedlo k zaujetí webscrapingem dat z Úřadu průmyslového vlastnictví. Proces zahrnoval opakované úpravy skriptu, což mě bavilo, a přitom mi poskytl cenné zkušenosti s Pythonem a s podporou ChatGPT. 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. Tato zkušenost mi ukázala, jak propojit technické dovednosti s vizualizací a komunikací výsledků.

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, což nám umožnilo sdílet výsledky projektu a naše získané poznatky.

Poděkování

Rády bychom poděkovaly celému týmu Czechitas, který se podílel na organizaci podzimního běhu Digitální Akademie Data v Brně. Naše díky patří především Martinovi Kovářovi, Adéle Bierské, Janči Prchlové, Milošovi Minaříkovi, Jirkovi Piškinovi, Anet Lukáčevičové, Tereze Fukátkové, Dii Fašangové, Zdeňce Herudkové a mnoha dalším. Za pomoc s projektem děkujeme mentorovi Honzovi Kotenovi, Adéle Bierské za rady s Regexem, Martinu Kovářovi za podporu na obou hackathonech a našemu kamarádovi Marianowi Kałużowi za asistenci s Gitem. 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

- 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, Googledocs

- Canva