



VYSOKÉ UČENÍ TECHNICKÉ V BRNĚ

BRNO UNIVERSITY OF TECHNOLOGY

FAKULTA PODNIKATELSKÁ

FACULTY OF BUSINESS AND MANAGEMENT

ÚSTAV INFORMATIKY

INSTITUTE OF INFORMATICS

METODIKA SPRÁVY DRG KÓDŮ: NÁVRH ŘEŠENÍ PRO KONTROLU DUPLICIT A ROČNÍ AKTUALIZACI

A STRUCTURED METHOD FOR DRG CODE MAINTENANCE: STRATEGIES FOR HANDLING DUPLICATES
AND YEARLY REVISIONS

BAKALÁŘSKÁ PRÁCE

BACHELOR'S THESIS

AUTOR PRÁCE

AUTHOR

Daniel Kopanica

VEDOUCÍ PRÁCE

SUPERVISOR

doc. Mgr. Veronika Novotná, Ph.D.

BRNO 2025

Zadání bakalářské práce

Ústav: Ústav informatiky
Student: **Daniel Kopanica**
Vedoucí práce: **doc. Mgr. Veronika Novotná, Ph.D.**
Akademický rok: 2024/25
Studijní program: Manažerská informatika

Garant studijního programu Vám v souladu se zákonem č. 111/1998 Sb., o vysokých školách ve znění pozdějších předpisů a se Studijním a zkušebním řádem VUT v Brně zadává bakalářskou práci s názvem:

Metodika správy DRG kódů: Návrh řešení pro kontrolu duplicit a roční aktualizaci

Charakteristika problematiky úkolu:

Úvod
Cíle práce, metody a postupy zpracování
Teoretická východiska práce
Analýza současného stavu
Vlastní návrhy řešení
Závěr
Seznam použité literatury
Přílohy

Cíle, kterých má být dosaženo:

Cílem práce je analyzovat změny v číselníku zdravotních výkonů a navrhnout řešení na odstranění zjištěných nedostatků a zabezpečení konzistence číselníku.

Základní literární prameny:

HOTEK, M. SQL Server 2008, krok za krokem. Brno, Computer Press, 2009, 488 s. ISBN 978-80-251-2466-6.

LACKO, Ľuboslav. Mistrovství v SQL Server 2012. Brno: Computer Press, 2013. ISBN 978-80-251-3773-4.

OPPEL, Andrew J. SQL bez předchozích znalostí: [průvodce pro samouky]. Brno: Computer Press, 2008. ISBN 978-80-251-1707-1.

PRATT, PHILIP J.; LAST, MARY Z. A Guide to SQL, Cengage Learning, Inc 2014, 360s., ISBN 9781111527273

STEPHENS, R.K. a R.R. PLEW. Naučte se SQL za 21 dní. Brno, Computer Press 2004, 491 s., ISBN 80-722-6870-8

Termín odevzdání bakalářské práce je stanoven časovým plánem akademického roku 2024/25

V Brně dne 9.2.2025

L. S.

Ing. Jiří Kříž, Ph.D.
garant

prof. Ing. et Ing. Stanislav Škapa, Ph.D.
děkan

Abstrakt

Táto bakalárska práca sa zaoberá analýzou číselníka zdravotníckych výkonov ZV-DRG a jeho spracovaním pre potreby importu do interného informačného systému Národného centra zdravotníckych informácií (NCZI). Vzhľadom na rozsiahlu aktualizáciu číselníka v roku 2024, ktorá spôsobila vznik duplícít, prekročenie povolenej dĺžky názvov a výskyt nekompatibilných znakov, bola vytvorená sada nástrojov na ich automatickú detekciu a opravu. Riešenie bolo implementované v jazyku Python s využitím knižníc pandas a openpyxl. Výsledkom je súbor skriptov, ktoré dokážu filtrovať relevantné záznamy, kontrolovať jednoznačnosť údajov, skrátiť dlhé názvy, odstrániť neželané znaky a vytvoriť výstupy pripravené na import vrátane verzovania. Práca poukazuje na výhody automatizácie a odporúča ďalšie smerovanie vývoja smerom k integrácii nástrojov do systémov NCZI.

Kľúčové slová

číselník, ZV-DRG, NCZI, Python, pandas, openpyxl, duplicita, automatizácia, zdravotnícky výkon, dátová analýza

Abstract

This bachelor's thesis focuses on the analysis of the ZV-DRG medical procedure code list and its processing for import into the internal information system of the National Health Information Centre (NCZI). Due to the extensive update of the code list in 2024, which led to the emergence of duplicates, overly long descriptions, and incompatible characters, a set of tools was developed for automatic detection and correction of these issues. The solution was implemented in Python using the pandas and openpyxl libraries. The result is a script collection that filters relevant records, checks data uniqueness, shortens long names, removes undesired characters, and creates versioned output files ready for import. The thesis highlights the benefits of automation and proposes further development toward integration with NCZI systems.

Keywords

code list, ZV-DRG, NCZI, Python, pandas, openpyxl, duplication, automation, medical procedure, data analysis

Bibliografická citace

KOPANICA, Daniel. *Metodika správy DRG kódů: Návrh řešení pro kontrolu duplicit a roční aktualizaci* [online]. Brno, 2025 [cit. 2025-05-15]. Dostupné z: <https://www.vutbr.cz/studenti/zav-prace/detail/168845>. Bakalářská práce. Vysoké učení technické v Brně, Fakulta podnikatelská, Ústav informatiky. Vedoucí práce doc. Mgr. Veronika Novotná, Ph.D.

Čestné prohlášení

Prohlašuji, že předložená bakalářská práce je původní a zpracoval jsem ji samostatně. Prohlašuji, že citace použitých pramenů je úplná, že jsem ve své práci neporušil autorská práva (ve smyslu zákona č. 121/2000 Sb., o právu autorském a o právech souvisejících s právem autorským).

V Brně dne 15. 5. 2025

Daniel Kopanica

autor

Pod'akovanie

Rád by som touto cestou poďakoval vedúcemu bakalárskej práce Ing. Jiřímu Křížovi, Ph.D. za pomoc, konzultácie a užitočné rady, ktoré mi poskytol pri písaní tejto bakalárskej práce a mojej mamine za psychickú pomoc a cenné rady.

Obsah

Úvod	11
Ciele práce a postupy spracovania.....	12
Ciele práce	12
Metódy spracovania	12
Postupy spracovania.....	12
1 Teoretické východiska riešenia.....	14
1.1 Základné pojmy	14
1.1.1 Informácia	14
1.1.2 Dáta	14
1.2 Databázy	14
1.2.1 Význam a základné pojmy	14
1.2.2 Primárny kľúč a funkčná závislosť	15
1.2.3 Význam správneho návrhu databázy	15
1.3 Číselníky v informačných systémoch	15
1.3.1 Význam číselníkov pri práci s údajmi.....	15
1.3.2 Definícia a štruktúra číselníka	16
1.4 Spracovanie dát pomocou jazyka Python	16
1.4.1 Typy dát vhodné na analýzu	16
1.4.2 Prečo práve Python?.....	16
1.4.5 Základné knižnice pre spracovanie dát v jazyku Python	17
2 Analýza súčasného stavu	18
2.1 Základné údaje o NCZI.....	18
2.2 Organizačná štruktúra NCZI.....	18
2.3 Technické zázemie NCZI(treba viac faktov).....	19

2.4 Informačné systémy a programy na podporu svojich činností v oblasti zdravotníctva	20
2.5 Problémy a výzvy, ktorým NCZI čelí.....	20
2.5.1 Práca s dátami a databázou v prostredí NCZI	21
2.5.2 Spracovanie číselníkov.....	21
2.5.3 Používané nástroje a automatizácia	21
2.5.4 Aktuálny stav spracovania dát	22
2.6 Výzvy do budúcnosti	22
2.7 Porovnanie spracovania dát v Pythone (pandas) a v Microsoft Excel.....	22
2.7.1 Spracovanie dát v Microsoft Excel.....	22
2.7.2 Spracovanie dát v Pythone	23
2.7.3 Vytváranie verzií súborov.....	23
3 Vlastný návrh riešenia	24
3.1 Požiadavky na vstupný súbor pre import do IIS	24
3.1.1 Filtrovanie relevantných záznamov	24
3.1.2 Kontrola jednoznačnosti kódov	24
3.1.3 Kontrola jednoznačnosti názvov výkonov	24
3.1.4 Odstránenie nekompatibilných znakov	24
3.1.5 Limit dĺžky názvu výkonu.....	25
3.2 Logický návrh riešenia problémov.....	26
3.2.1 Filtrovanie relevantných záznamov	26
3.2.2 Kontrola duplicit v kódoch a názvoch	27
3.2.3 Odstránenie nekompatibilných znakov	28
3.2.4 Orezanie názvu na 128 znakov	29
3.2.5 Porovnanie číselníkov 2024 vs. 2025 (detekcia nových kódov)	30
3.2.6 Zápis výstupných súborov (export, verzionovanie).....	31

3.3 Vlastné riešenie	32
3.3.1 vybrať len tie riadky, kde hodnoty prvého stĺpca = “T“	33
3.3.2 Kontrola jednoznačnosti kódov a názvov v číselníku ZV-DRG	34
3.3.3 odstrániť znak „.“ z kódu číselníka	40
3.3.4 počet znakov, ktorý má Názov ZV-DRG 2024 nesmie presiahnuť 128.....	42
3.3.5 Aktualizácia číselníka na rok 2025	46
3.4 Význam implementácie skriptu v praxi a odporúčania pre NCZI	52
Záver	53
Zoznam použitej literatúry	55
Zoznam použitých obrázkov	56
Zoznam použitých tabuliek	57

Úvod

Správne spracovanie a aktualizácia číselníkov je kľúčovým aspektom efektívneho fungovania zdravotníckych systémov. Číselníky zdravotníckych výkonov pre DRG (Diagnosis-Related Groups) slúžia ako základný nástroj pre klasifikáciu, fakturáciu a analýzu zdravotníckych služieb.

V roku 2024 došlo k rozsiahlej aktualizácii číselníka DRG, ktorá priniesla nárast o približne 16 000 nových kódov. Táto aktualizácia však nespĺňa stanovené pravidlá tvorby číselníka, čo spôsobilo vznik rôznych nejasností, ako sú duplicity kódov a popisov. Tieto nezrovnalosti môžu negatívne ovplyvniť použiteľnosť číselníka a spôsobiť problémy pri jeho implementácii do praxe.

Táto bakalárska práca sa zameriava na analýzu zmien v číselníku zdravotníckych výkonov pre DRG s cieľom identifikovať problematické oblasti, navrhnúť riešenia na ich odstránenie a prispieť k optimalizácii jeho správy.

Ciele práce a postupy spracovania

Ciele práce

Cieľom práce je analyzovať zmeny v číselníku zdravotníckych výkonov a navrhnúť riešenia na odstránenie zistených nedostatkov a zabezpečenie konzistencie číselníka.:

Metódy spracovania

Pri spracovaní tejto bakalárskej práce budú uplatnené metódy, ktoré umožnia systematickú analýzu zmien v číselníkoch zdravotníckych výkonov.

V úvodnej etape bude využitá literárna rešerš, ktorá pomôže získať základné poznatky o problematike číselníkov, ich štruktúre, správe a význame v informačných systémoch.

Na základe týchto poznatkov bude vytvorený koncept práce s dátami.

Ďalej sa predpokladá využitie porovnávacjej analýzy dát, pri ktorej budú vyhľadávané zmeny medzi číselníkmi v dvoch po sebe nasledujúcich obdobiach. Táto analýza môže byť realizovaná buď pomocou tabuľkových nástrojov (napr. Microsoft Excel), alebo pomocou skriptovacích jazykov (napr. Python), ktoré umožňujú automatizáciu úloh, efektívnejšiu manipuláciu s údajmi a generovanie výstupov. Konkrétny výber nástroja bude závisieť od rozsahu dát a požiadaviek na presnosť a opakovateľnosť.

Záverom bude vykonaná interpretácia výsledkov s cieľom identifikovať nedostatky v číselníkoch a navrhnúť možné riešenia pre zvýšenie konzistencie, jednoznačnosti a správnosti údajov.

Postupy spracovania

Postup spracovania tejto bakalárskej práce bude pozostávať z viacerých naväzujúcich krokov, ktorých cieľom je zabezpečiť systematickú analýzu číselníkov zdravotných výkonov. V prvej fáze bude potrebné zhromaždiť relevantné dátové súbory číselníkov za roky 2024 a 2025, ktoré budú slúžiť ako vstup pre ďalšie spracovanie. Následne bude vykonaná analýza procesov a úloh, ktoré bude potrebné realizovať na dosiahnutie stanovených cieľov – najmä identifikácia rozdielov, nových a neplatných kódov, ako aj prípadných duplicit.

Na základe zistení bude navrhnutý spôsob technickej implementácie porovnávacjej analýzy. V tomto kroku sa predpokladá využitie vhodného programovacieho alebo

analytického nástroja, pričom konkrétny výber technológie bude určený podľa povahy údajov a požiadaviek na výstupy. Po vykonaní analýzy budú výsledky vyhodnotené a interpretované s dôrazom na ich praktickú využiteľnosť. Výstupy budú slúžiť ako podklad pre ďalšie návrhy riešení, ktoré budú predmetom samostatnej kapitoly práce.

1 Teoretické východiska riešenia

1.1 Základné pojmy

1.1.1 Informácia

Informáciou rozumieme správu alebo vnem, ktorý musí zároveň spĺňať tri základné požiadavky:

- Syntaktická relevantnosť – prijímateľ správy musí byť schopný túto správu rozpoznať, teda detekovať jej štruktúru.
- Sémantická relevantnosť – prijímateľ musí správu pochopiť, teda rozumieť obsahu správy.
- Pragmatická relevantnosť – informácia musí mať pre prijímateľa význam, teda musí byť v danom kontexte použiteľná.

Informácia vzniká až vtedy, keď prijímateľ správu nielen zachytí, ale jej aj porozumie a priradí jej význam.

Informačný tok preto nie je len technickým prenosom údajov, ale procesom, v ktorom je zmysel správy aktívne vytváraný zo strany prijímateľa (1).

1.1.2 Dáta

Dáta predstavujú správy alebo záznamy, ktoré môžu niesť význam – sú teda potenciálnymi informáciami. Ak subjekt dáta používa na rozhodovanie, priraduje im význam, a tým sa stávajú informáciou .

Dáta možno zaznamenávať na rôzne fyzické médiá, ako je papier, magnetické alebo elektronické nosiče. Tieto zaznamenané dáta môžu byť uchovávané pre neskoršie spracovanie a po ich prečítaní (dekódovaní) môžu opätovne nadobudnúť formu informácie. Ide teda o reverzibilný proces, v ktorom sa informácia mení na dáta a dáta opäť na informáciu – podľa kontextu použitia (1).

1.2 Databázy

1.2.1 Význam a základné pojmy

Databáza je pojem, ktorý je dnes neoddeliteľnou súčasťou fungovania modernej spoločnosti. Systémy na ukladanie a správu údajov sa využívajú v zdravotníctve, pri evidencii obyvateľstva, vo vede, priemysle, školstve či telekomunikáciách.

Umožňujú systematické zhromažďovanie, uchovávanie a spracovanie informácií, ktoré by inak boli roztriešené alebo ťažko dostupné (2).

Jednou z najdôležitejších databázových štruktúr je relačný model, kde sa údaje ukladajú do tabuliek – relácií. Každá relácia predstavuje jednu entitu (napr. zamestnanec) a jej atribúty (napr. meno, dátum narodenia, plat). Riadok v tabuľke predstavuje jeden záznam – teda jednu konkrétnu inštanciu tejto entity (2).

1.2.2 Primárny kľúč a funkčná závislosť

Každý stĺpec v databázovej tabuľke má svoj dátový typ (napr. celé číslo, reťazec, dátum). Z hľadiska integrity údajov je dôležité, aby tabuľka mala jeden alebo viac primárnych kľúčov – teda takých atribútov, ktorých kombinácia jednoznačne identifikuje každý záznam. Príkladom môže byť osobné číslo zamestnanca (2).

S pojmom primárneho kľúča úzko súvisí aj funkčná závislosť – teda vzťah, v ktorom hodnota jednej položky závisí od hodnoty inej. Napríklad, ak poznáme pracovnú pozíciu, môžeme predpokladať platovú kategóriu. Takéto väzby sa často označujú symbolicky, napríklad: FUNKCIA → PLAT alebo DĹŽKA_TRASY → VÝŠKA_JAZDNÉHO (2).

1.2.3 Význam správneho návrhu databázy

Dobre navrhnutá databáza je základom pre spoľahlivý a efektívny informačný systém. Nesprávne navrhnutá tabuľková štruktúra môže viesť k duplicitám, nekonzistentnosti údajov a zníženej výkonnosti systému. Preto sa v praxi uplatňujú tzv. normálne formy, ktoré zabezpečujú, že tabuľky spĺňajú určité logické a technické kritériá pre správne usporiadanie údajov (2).

1.3 Číselníky v informačných systémoch

1.3.1 Význam číselníkov pri práci s údajmi

Číselníky sú nástroje na štandardizáciu údajov v informačných systémoch. Ich cieľom je zabezpečiť, aby boli údaje zaznamenávané a spracovávané jednotne, bez nejednoznačností. To je kľúčové najmä v prostrediach, kde sa pracuje s veľkým objemom štruktúrovaných údajov, ako napríklad v zdravotníctve, ekonomike alebo verejnej správe (1).

Vďaka číselníkom sa minimalizuje riziko nekonzistencie údajov a zvyšuje sa efektivita ich spracovania. Napríklad namiesto zadávania voľného textu pre diagnózu sa použije štandardizovaný kód, ktorý jednoznačne určuje, o aký zdravotný stav ide (1).

1.3.2 Definícia a štruktúra číselníka

Číselník možno definovať ako formálne definovaný a centrálny spravovaný zoznam hodnôt, ktoré majú presne stanovený význam. Každý záznam v číselníku typicky obsahuje kód, názov položky, dátum platnosti, stav (napr. aktívny/neplatný) a prípadné poznámky (4).

Tieto záznamy sú často organizované hierarchicky, čo znamená, že jeden kód môže byť nadradený viacerým podkódom. Táto štruktúra umožňuje systematickú klasifikáciu veľkého množstva údajov v zložitých doménach, ako je klinická medicína (4).

1.4 Spracovanie dát pomocou jazyka Python

1.4.1 Typy dát vhodné na analýzu

Pri práci s dátami sa v praxi stretávame najčastejšie so štruktúrovanými údajmi, ktoré sú vhodné na spracovanie pomocou knižníc ako pandas či numpy. Medzi najbežnejšie formáty patria tabuľkové údaje (napr. Excel, CSV), viacrozmerné polia (matice), prepojené tabuľky (cez primárne a cudzie kľúče) a časové rady – pravidelné aj nepravidelné (3).

Dôležité je, že aj nestruktúrované dáta (napr. texty článkov) môžu byť transformované do štruktúrovanej formy, napríklad do frekvenčnej tabuľky slov, ktorú je následne možné analyzovať alebo modelovať (3).

1.4.2 Prečo práve Python?

Python je interpretovaný programovací jazyk, ktorý sa od roku 2000 stal jedným z najpoužívanejších nástrojov na vedecké výpočty a dátovú analýzu. Vďaka veľkej komunite a knižniciam ako pandas, numpy, matplotlib, scikit-learn či openpyxl, je Python vhodný nielen na prototypovanie, ale aj na tvorbu produkčných systémov (3).

V porovnaní s inými nástrojmi ako R, MATLAB či SAS má Python výhodu v tom, že prepája analytické schopnosti s možnosťou vývoja plnohodnotných softvérových riešení. Vďaka tomu možno v rámci jedného ekosystému prejsť od spracovania dát až po implementáciu výstupu do aplikácie (3).

1.4.5 Základné knižnice pre spracovanie dát v jazyku Python

Pri spracovaní štruktúrovaných údajov v prostredí jazyka Python zohrávajú kľúčovú úlohu dve knižnice: NumPy a pandas. Obe tvoria základ dátovej analytiky a umožňujú efektívnu manipuláciu s tabuľkovými, maticovými či časovo indexovanými dátami (3).

1.4.5.1 NumPy

Knižnica NumPy (Numerical Python) poskytuje výkonný objekt `ndarray`, čo je viacrozmerné pole, ktoré umožňuje efektívne uchovávanie a výpočty s numerickými údajmi. Okrem toho obsahuje funkcie pre lineárnu algebru, generovanie náhodných čísel, Fourierovu transformáciu či rýchle čítanie a zapisovanie binárnych súborov s maticovými dátami (3).

Jednou z hlavných výhod NumPy je schopnosť interoperovať s knižnicami v jazykoch C a Fortran bez nutnosti kopírovania dát, čím sa výrazne zvyšuje výkon. Preto je `ndarray` často používaný ako základný kontajner pre numerické údaje, ktoré sa odovzdávajú medzi algoritmami a knižnicami (3).

1.4.5.2 Pandas

Knižnica Pandas poskytuje vyššie úrovne dátových štruktúr – najmä objekt `DataFrame`, ktorý umožňuje prácu s tabuľkovými údajmi s označenými riadkami a stĺpcami, a objekt `Series`, čo je jednorozmerné pole s menovkami. Pandas bol navrhnutý s cieľom spojiť výhody výkonnej maticovej aritmetiky z NumPy s flexibilitou databázového a tabuľkového spracovania (3).

2 Analýza súčasného stavu

2.1 Základné údaje o NCZI

Národné centrum zdravotníckych informácií (NCZI) je kľúčovou inštitúciou v oblasti zdravotníctva na Slovensku, ktorá zohráva významnú úlohu pri zhromažďovaní, spracovávaní a analýze zdravotníckych údajov. Jeho hlavné poslanie spočíva v poskytovaní relevantných informácií a dát, ktoré sú potrebné na podporu rozhodovacích procesov v rámci zdravotníckej politiky a plánovania.

Medzi základné činnosti NCZI patrí správa zdravotníckych informačných systémov, ako je napríklad eZdravie, ktoré umožňuje elektronickú komunikáciu medzi zdravotníckymi zariadeniami a pacientmi, či elektronická zdravotná karta, ktorá zhromažďuje údaje o zdravotnom stave pacientov. NCZI taktiež spravuje viaceré národné registre, ako napríklad Národný onkologický register.

Jednou z hlavných priorít NCZI je digitalizácia zdravotníctva. Táto inštitúcia sa zameriava na zavádzanie moderných technologických riešení, ktoré umožňujú efektívnejšiu komunikáciu a správu zdravotníckych informácií. Zároveň kladie dôraz na ochranu týchto údajov, čím zabezpečuje ich bezpečnosť pred neoprávneným prístupom alebo stratou.

Hlavnými klientmi NCZI sú štátne inštitúcie, zdravotnícke zariadenia, zdravotné poisťovne a samotná verejnosť. Patrí pod Ministerstvo zdravotníctva Slovenskej republiky a spolupracuje s ďalšími organizáciami, ktoré pôsobia v oblasti zdravotníctva, aby zabezpečilo efektívnu koordináciu a využitie dostupných zdrojov.

2.2 Organizačná štruktúra NCZI.

NCZI má dobre definovanú organizačnú štruktúru. Táto štruktúra je rozdelená na viacero oddelení, z ktorých každé má svoju špecifickú zodpovednosť. Na čele organizácie stojí generálny riaditeľ, ktorý zodpovedá za strategické rozhodnutia a smerovanie centra. Pri jeho práci mu asistuje manažérsky tím, ktorý koordinuje jednotlivé oddelenia.

Jedným z kľúčových oddelení je IT sekcia, ktorá sa zaoberá technickou správou systémov, vývojom softvéru a údržbou databáz. Toto oddelenie zabezpečuje technickú infraštruktúru potrebnú na efektívne spracovanie zdravotníckych dát a prevádzku digitálnych nástrojov, ako je eZdravie. Významnú rolu zohráva aj oddelenie analýzy dát,

ktoré spracováva štatistické a analytické výstupy potrebné pre rozhodovanie v oblasti zdravotnej politiky.

Ďalšou dôležitou súčasťou organizačnej štruktúry je oddelenie pre legislatívu a compliance, ktoré dohliada na to, aby všetky činnosti NCZI spĺňali platné právne predpisy, najmä v oblasti ochrany osobných údajov a zdravotníckej dokumentácie. Taktiež sa zameriava na implementáciu legislatívnych zmien do praxe.

Z pohľadu vnútornej organizácie je dôležité aj oddelenie projektového manažmentu, ktoré riadi kľúčové projekty NCZI a zabezpečuje ich súlad s celkovými cieľmi organizácie. Na dennej báze spolupracuje s ostatnými oddeleniami, aby zabezpečilo efektívne vykonávanie jednotlivých úloh.

2.3 Technické zázemie NCZI

Ako kľúčová inštitúcia pre spracovanie a správu zdravotníckych dát, NCZI využíva moderné informačné technológie a databázové systémy, ktoré umožňujú spracovanie veľkého množstva údajov a ich bezpečné uchovávanie.

Databázové systémy, ktoré NCZI používa, sú optimalizované pre spracovanie štruktúrovaných údajov. Jedným z najpoužívanejších nástrojov je SQL, ktorý umožňuje efektívne spracovanie dotazov, analýzu dát a vytváranie štatistických výstupov. Tento nástroj je obzvlášť dôležitý pri správe číselníkov zdravotníckych výkonov a ich aktualizácií.

Bezpečnosť dát je pre NCZI prioritou. Sú implementované pokročilé bezpečnostné mechanizmy, ktoré chránia zdravotnícke údaje pred neoprávneným prístupom alebo únikom. Tieto mechanizmy zahŕňajú šifrovanie údajov, viacfaktorové overovanie používateľov a pravidelné audity systémov.

Okrem toho NCZI neustále investuje do technologických inovácií. Príkladom je zavádzanie analytických nástrojov na báze umelej inteligencie, ktoré umožňujú pokročilé spracovanie dát a prediktívne modelovanie. Tieto technológie môžu výrazne zlepšiť kvalitu rozhodovania a efektívnosť zdravotníckeho systému.

2.4 Informačné systémy a programy na podporu svojich činností v oblasti zdravotníctva

eZdravie: Komplexný systém elektronického zdravotníctva, ktorý zahŕňa elektronickú zdravotnú knižku pacienta, eRecept a ďalšie služby umožňujúce digitálnu komunikáciu medzi poskytovateľmi zdravotnej starostlivosti a pacientmi.

Informačný systém zdravotníckych indikátorov (ISZI): Systém zameraný na zber, spracovanie a poskytovanie zdravotníckych indikátorov a údajov o zdravotníctve v Slovenskej republike. Umožňuje automatizovanú podporu procesov, ako je návrh zisťovania, zber údajov a výstup údajov.

Národné zdravotné registre: Špecifické informačné systémy zhromažďujúce, spracúvajúce a analyzujúce údaje o hromadne sa vyskytujúcich a spoločensky závažných skupinách chorôb v SR. Tieto registre monitorujú vývoj počtu novo diagnostikovaných pacientov a podporujú optimalizáciu intervenčných opatrení v zdravotníctve.

Národný zdravotný portál (NZZP): Jednotné miesto na internete pre prístup k zdravotníckym informáciám pre laickú aj odbornú verejnosť, vrátane účastníkov zdravotnej starostlivosti. NZZP spĺňa relevantné štandardy prístupnosti a funkčnosti.

Číselníky a klasifikácie: NCZI spravuje rôzne číselníky a klasifikácie pre zdravotnícke informačné systémy, ako napríklad SNOMED CT, ICPC-2 a SNOLAMED, ktoré zabezpečujú štandardizáciu v oblasti zdravotníckej informatiky.

2.5 Problémy a výzvy, ktorým NCZI čelí.

Jedným z najvýraznejších problémov je zastaraná technická infraštruktúra. Napriek pokrokom v implementácii systémov, mnohé technológie používané v NCZI vyžadujú modernizáciu, aby zvládali rastúce nároky na spracovanie zdravotníckych údajov. Nedostatočne výkonné systémy môžu spôsobiť oneskorenia v spracovaní dát alebo problémy s dostupnosťou služieb.

Ďalšou výzvou je kvalita a konzistentnosť zdravotníckych údajov. Kvôli rozdielom v reportovacích praktikách jednotlivých zdravotníckych zariadení dochádza k nezrovnalostiam v záznamoch, čo komplikuje analýzu a využitie dát. Tento problém je obzvlášť významný v súvislosti so spravovaním číselníkov a klasifikácií, kde sú presnosť a jednotnosť kľúčové pre správnu interpretáciu údajov.

NCZI taktiež čelí výzvam v oblasti kybernetickej bezpečnosti. Ako inštitúcia, ktorá spravuje citlivé zdravotnícke údaje, musí čeliť neustálym hrozbám neoprávneného prístupu a potenciálnych útokov na systémy. Zabezpečenie údajov si vyžaduje nielen pravidelné aktualizácie softvéru a hardvéru, ale aj školenia zamestnancov a implementáciu najmodernejších bezpečnostných štandardov.

V neposlednom rade je výzvou aj adaptácia na legislatívne zmeny a nové regulačné požiadavky. Systémy NCZI musia byť neustále aktualizované, aby spĺňali platné predpisy a štandardy, čo si vyžaduje značné finančné a časové zdroje.

2.5.1 Práca s dátami a databázou v prostredí NCZI

V rámci správy číselníkov zdravotníckych výkonov sa v NCZI využívajú štruktúrované dátové súbory, ktoré obsahujú údaje o zdravotníckych výkonoch vrátane ich kódov, popisov a platnosti. Tieto dáta sú spracovávané v databázových systémoch, kde hlavným cieľom je zabezpečiť ich konzistentnosť, jednoznačnosť a aktuálnosť.

Prístup k databázam majú predovšetkým pracovníci IT oddelenia, zatiaľ čo ostatní používatelia pracujú s dátami prostredníctvom aplikácií alebo formulárov.

2.5.2 Spracovanie číselníkov

Číselníky sú zvyčajne dostupné ako súbory vo formáte Excel, ktoré sú neskôr importované do centrálného systému. Na ich spracovanie sa využívajú rôzne softvérové nástroje – či už priamo SQL, alebo externé nástroje. V tejto práci som sa rozhodol pre druhý prístup – externé spracovanie číselníkov s cieľom odhaliť duplicity, nejednoznačnosti a nové kódy, ktoré sa v predchádzajúcich ročníkoch nevyskytovali.

2.5.3 Používané nástroje a automatizácia

Python ako skriptovací jazyk poskytuje možnosť vytvárať vlastné nástroje, ktoré automatizujú opakujúce sa úlohy. V mojom riešení som využil:

- knižnicu pandas na načítanie a manipuláciu s tabuľkovými dátami,
- knižnicu openpyxl na zápis do Excelu a zvýrazňovanie duplicit,
- vlastné funkcie pre verzovanie súborov, detekciu zmien a spracovanie textových obmedzení (napr. maximálna dĺžka názvu).

Tieto nástroje umožnili efektívne spracovanie stoviek riadkov údajov a vytvorenie prehľadných výstupov, ktoré sú pripravené na ďalší import do databázových systémov.

2.5.4 Aktuálny stav spracovania dát

Pri analýze číselníkov za roky 2024 a 2025 som identifikoval viaceré nedostatkov:

- Duplicity v kódoch alebo názvoch výkonov.
- Názvy presahujúce maximálnu dĺžku 128 znakov, ktoré môžu spôsobovať chyby pri importe.
- Prítomnosť znaku „“ v kódoch, ktorý je v niektorých systémoch nežiaduci.

Tieto problémy boli riešené automatizovanými skriptmi, ktoré dokážu nielen tieto chyby odhaliť, ale aj navrhnúť ich opravu (napr. skrátenie názvu, zlúčenie informácií o kóde a názve). Súčasťou riešenia je aj analýza nových kódov, ktoré sa v novom číselníku objavili prvýkrát.

2.6 Výzvy do budúcnosti

Riešenie implementované v tejto práci ukazuje, že aj jednoduché automatizačné nástroje môžu výrazne zefektívniť prácu s číselníkmi. Do budúcnosti by bolo vhodné integrovať podobné kontroly priamo do prostredia NCZI – napríklad vo forme validácie pred importom alebo pomocou webových rozhraní s upozoreniami. Takéto nástroje by zvýšili kvalitu údajov a znížili chybovosť pri práci s rozsiahlymi dátovými súbormi.

2.7 Porovnanie spracovania dát v Pythone (pandas) a v Microsoft Excel

Pri spracovaní rozsiahlych dátových súborov, ako sú číselníky zdravotných výkonov, je dôležité zvoliť efektívny nástroj, ktorý umožňuje rýchlu a spoľahlivú manipuláciu s údajmi. V rámci tejto práce bolo porovnávané spracovanie údajov v prostredí Microsoft Excel a prostredí Python, konkrétne s využitím knižnice pandas.

2.7.1 Spracovanie dát v Microsoft Excel

Spracovanie dát v prostredí Microsoft Excel ponúka základné nástroje, ako sú filtrovanie, vyhľadávanie duplicit, odstraňovanie nežiaducich znakov alebo spájanie tabuliek. Pri práci s menšími súbormi sa Excel javí ako dostatočné riešenie. S narastajúcim objemom údajov sa však prejavujú jeho viaceré nevýhody. Ručné spracovanie je časovo náročné a zároveň náchylné na vznik chýb.

Navyše, pri práci s veľkými tabuľkami môže dôjsť k výraznému spomaleniu – operácie prebiehajú pomaly a načítavanie súborov trvá dlhšie. Verziovanie dokumentov nie je automatizované; používateľ musí staršiu verziu uložiť manuálne pod iným názvom.

Pri zložitejších úlohách je zároveň potrebné využívať makrá alebo vykonávať manuálne zásahy, čo obmedzuje možnosť plnej automatizácie.

2.7.2 Spracovanie dát v Pythone

Programovací jazyk Python spolu s knižnicou pandas predstavuje efektívnejší prístup k spracovaniu dát, najmä v prípadoch, keď ide o väčší rozsah údajov alebo opakovateľné úlohy. Pandas umožňuje načítavať, filtrovať, upravovať a exportovať státisíce riadkov v priebehu niekoľkých sekúnd. Pomocou vlastných funkcií možno celý proces plne automatizovať – či už ide o filtrovanie dát podľa zvolených kritérií (napr. typ výkonu), identifikáciu a označovanie duplicitných záznamov, úpravu názvov presahujúcich definovanú dĺžku alebo generovanie nových verzií súborov pri každom uložení. Navyše, napísaný skript je možné opakovane použiť na nové dáta bez potreby dodatočných manuálnych úprav, čo výrazne zvyšuje efektívnosť a škálovateľnosť riešenia.

2.7.3 Vytváranie verzií súborov

Významnou výhodou využitia Pythonu bolo vytvorenie vlastnej funkcie pre automatické verzionovanie výstupov. Táto funkcia zabezpečuje, že každý exportovaný súbor má v názve číslo verzie (napr. `Import_2024_modified_v1.xlsx`, `Import_2024_modified_v2.xlsx`), čo uľahčuje kontrolu zmien a zamedzuje prepisovaniu existujúcich údajov.

3 Vlastný návrh riešenia

Hlavným cieľom bakalárskej práce je analyzovať zmeny v číselníku zdravotníckych výkonov DRG medzi rokmi 2024 a 2025. Cieľom je nielen identifikovať zmeny, ale aj navrhnúť spôsoby, ako zabezpečiť vyššiu konzistenciu a jednoznačnosť údajov v číselníku. Tento hlavný cieľ sa ďalej rozčleňuje na požiadavky od Zadávateľa :

3.1 Požiadavky na vstupný súbor pre import do IIS

V tejto časti sú uvedené konkrétne technické požiadavky, ktoré musí spĺňať vstupný súbor, aby mohol byť úspešne naimportovaný do interného informačného systému (IIS). Tieto požiadavky zabezpečujú konzistentnosť, jednoznačnosť a technickú vhodnosť údajov.

3.1.1 Filtrovanie relevantných záznamov

Pre potreby importu je potrebné vybrať len tie riadky, ktoré reprezentujú najnižšiu úroveň hierarchie číselníka. Túto úroveň označuje hodnota „T“ v stĺpci Typ ZV-DRG. Ostatné záznamy s hodnotami ako „1M“, „2M“ atď. majú len orientačný alebo skupinový charakter a nie sú určené na priame použitie v systéme.

3.1.2 Kontrola jednoznačnosti kódov

Každý zdravotný výkon musí byť jednoznačne identifikovateľný pomocou unikátneho kódu (Kód ZV-DRG). Duplicitné výskyty toho istého kódu by mohli spôsobiť nejednoznačné spracovanie údajov, a preto je nevyhnutné, aby boli všetky kódy v rámci súboru jedinečné.

3.1.3 Kontrola jednoznačnosti názvov výkonov

Rovnako ako pri kódoch, aj názvy výkonov musia byť v rámci číselníka jednoznačné. V prípade, že rôzne kódy zdieľajú rovnaký názov, môže dôjsť k nesprávnemu zlúčeniu alebo interpretácii údajov v analytických výstupoch. Táto podmienka je obzvlášť dôležitá pri triedení, filtrovaní a zoskupovaní údajov.

3.1.4 Odstránenie nekompatibilných znakov

Interný systém IIS vyžaduje, aby kódy výkonov neobsahovali špeciálne znaky, ako napríklad bodku („.“). Preto je potrebné z každého kódu tento znak odstrániť. V opačnom prípade môže dôjsť k nesprávnemu zobrazeniu alebo interpretácii údajov – najmä

pri exporte do Excelu, kde sa niektoré hodnoty môžu automaticky konvertovať na vedecký formát (napr. 3.51E+05).

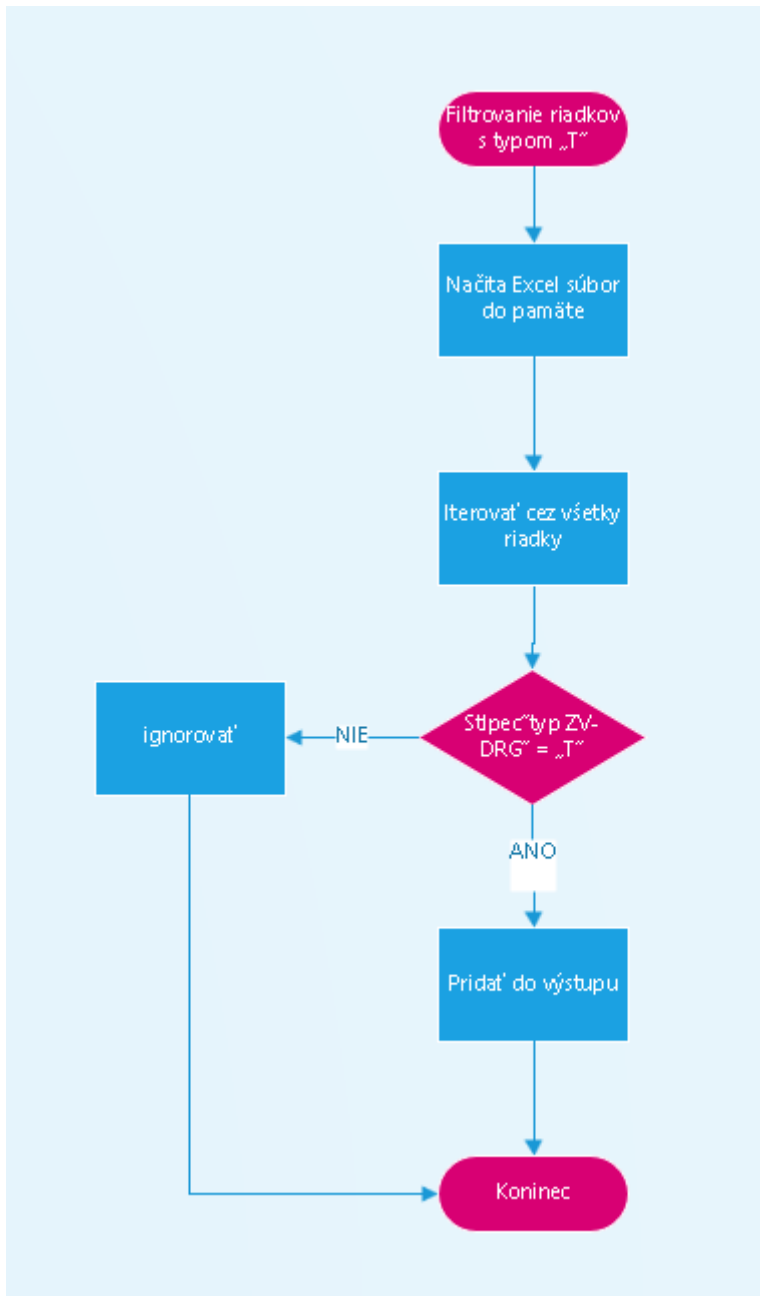
3.1.5 Limit dĺžky názvu výkonu

Maximálna povolená dĺžka názvu výkonu v systéme je 128 znakov. Pokiaľ pôvodný názov túto hranicu prekračuje, je potrebné ho skrátiť. Na zabezpečenie jednoznačnosti sa často aplikuje technika, pri ktorej sa k názvu pripojí aj jeho kód, a následne sa výsledný reťazec skráti na maximálnu povolenú dĺžku. Tým sa zároveň zaručí, že ani po skrátaní nedôjde k vzniku duplicitných názvov.

3.2 Logický návrh riešenia problémov

3.2.1 Filtrovanie relevantných záznamov

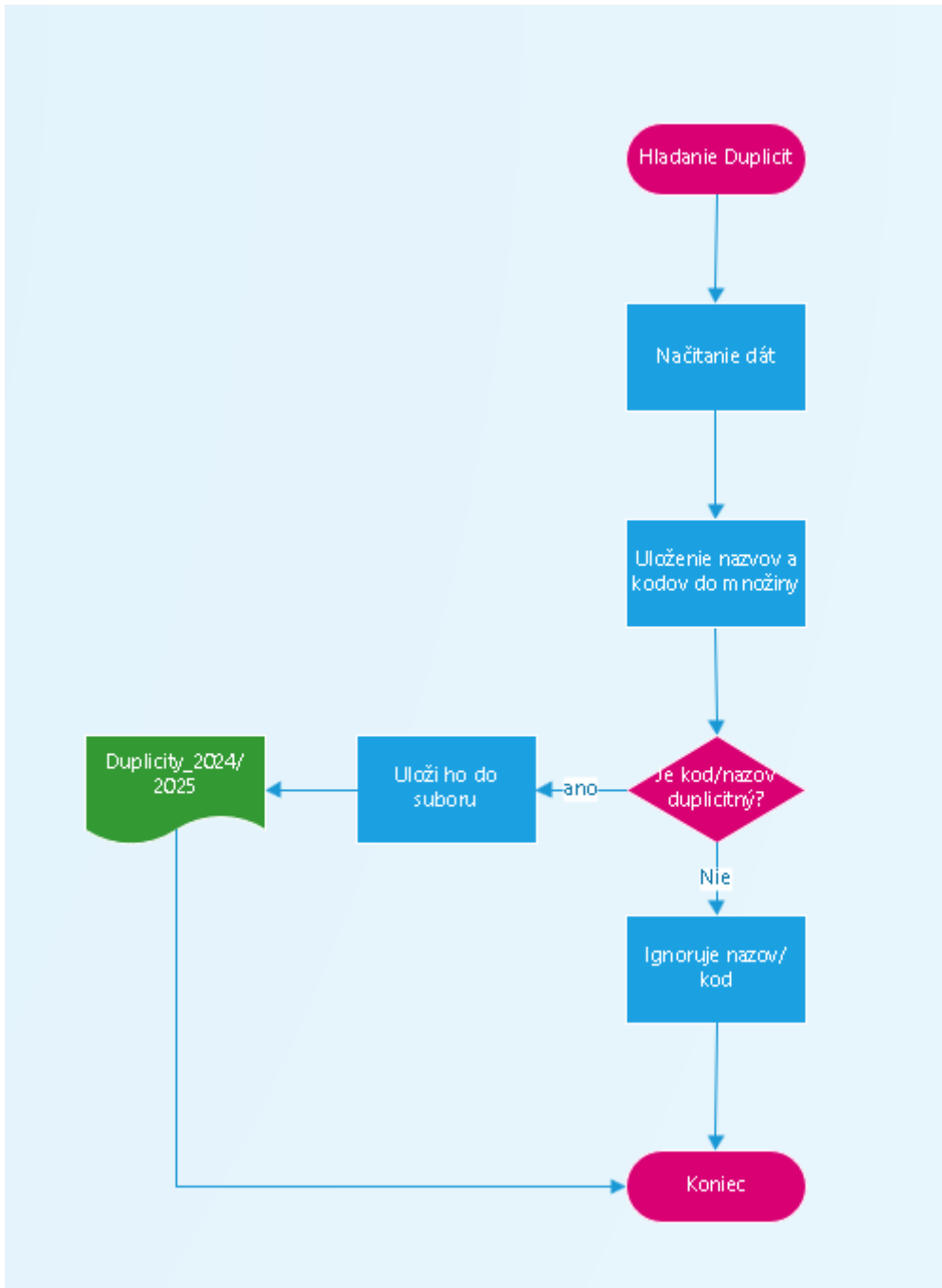
Znázornený logický návrh algoritmu, ktorý načíta Excel súbor, prechádza všetky jeho riadky a na základe hodnoty v stĺpci „typ ZV-DRG“ vyfiltruje iba tie záznamy, kde je táto hodnota rovná „T“. Tieto riadky sú následne pridané do výstupu, ostatné sú ignorované. Proces končí po spracovaní všetkých riadkov.



Obrázok 1: Filtrovanie relevantných záznamov (Vlastné)

3.2.2 Kontrola duplicit v kódoch a názvoch

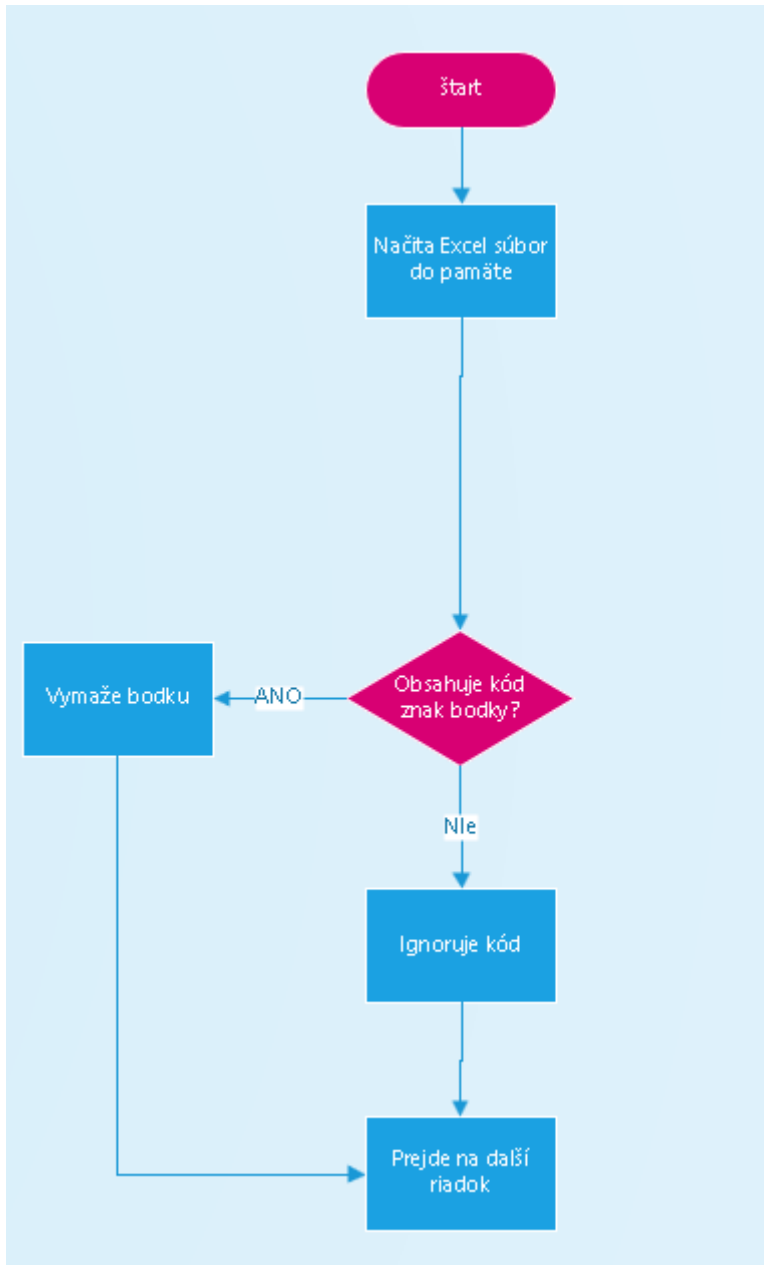
Diagram začína načítaním dát a následným uložením názvov a kódov do množiny. V ďalšom kroku sa overuje, či je daný kód alebo názov duplicitný. Ak áno, je pridaný do súboru s výsledkami (napr. „Duplicity_2024_2025“) a uložený. Ak duplicitný nie je, položka je ignorovaná. Celý proces pokračuje, kým nie sú spracované všetky dáta, a následne sa ukončí.



Obrázok 2: Kontrola duplicit v kódoch a názvoch (Vlastné)

3.2.3 Odstránenie nekompatibilných znakov

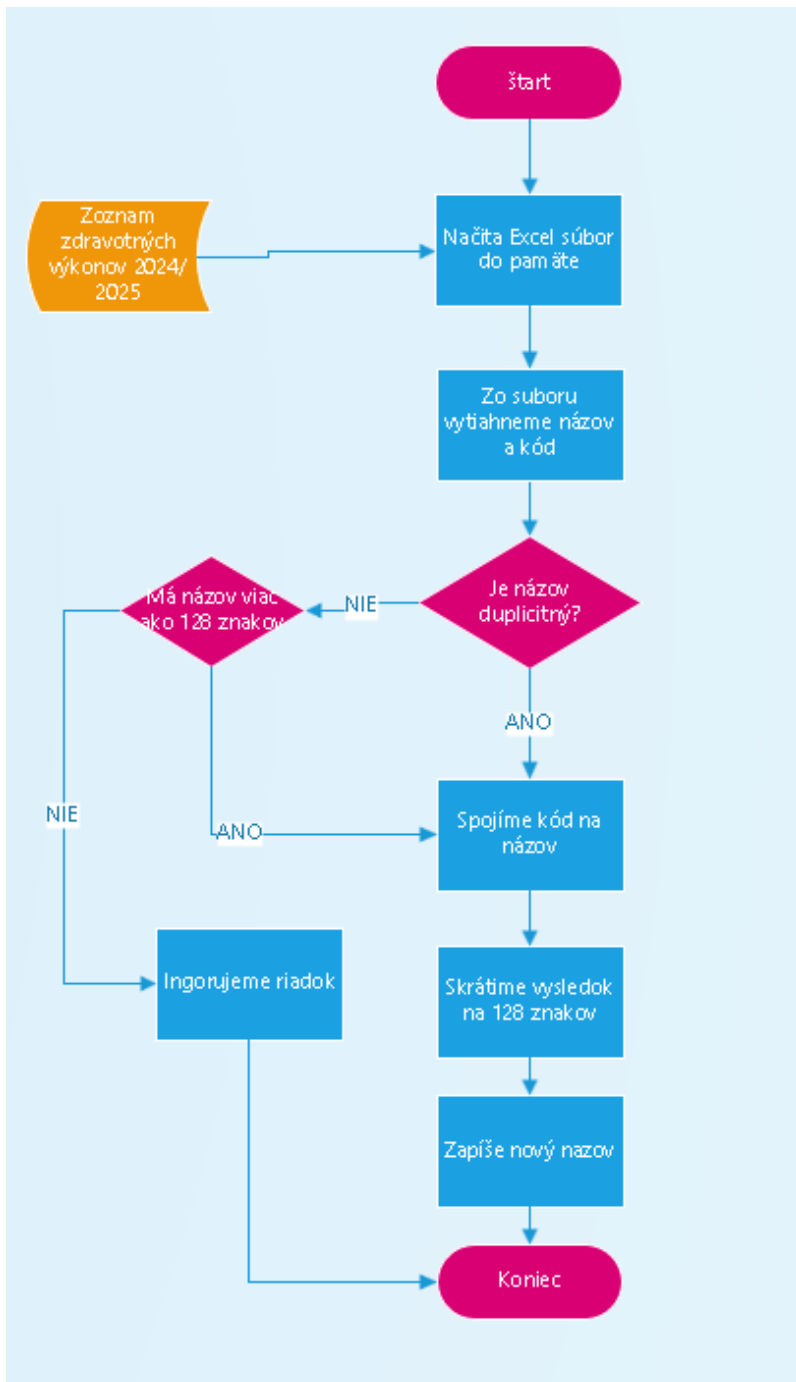
Diagram znázorňuje postup pri spracovaní Excel súboru s cieľom odstrániť bodky zo zdravotníckych kódov. Po načítaní súboru do pamäte sa každý riadok analyzuje. Ak kód obsahuje znak bodky, odstráni sa, v opačnom prípade sa kód ignoruje. Po vyhodnotení každého kódu, algoritmus pokračuje spracovaním ďalšieho riadku, až kým nie sú skontrolované všetky položky.



Obrázok 3: Odstránenie nekompatibilných znakov (Vlastné)

3.2.4 Orezanie názvu na 128 znakov

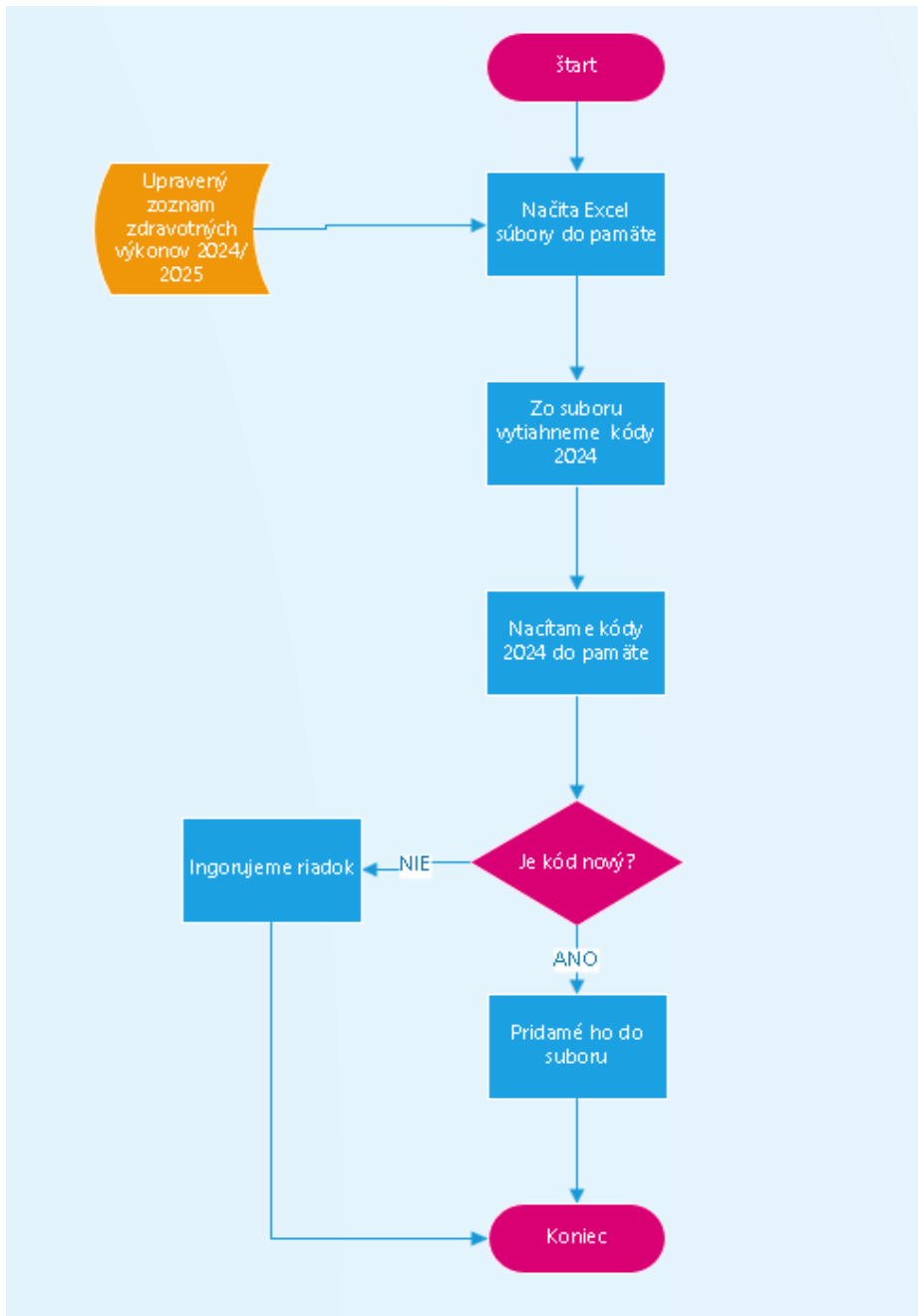
Najprv sa načíta Excel súbor do pamäte a z neho sa extrahujú názvy a kódy. Následne sa vykoná kontrola, či názov presahuje 128 znakov alebo je duplicitný. Ak je názov dlhý alebo sa opakuje, skombinuje sa s kódom, výsledok sa skráti na povolenú dĺžku a prepíše sa do výstupu. Ak tieto podmienky nie sú splnené, riadok sa ignoruje. Proces končí po spracovaní všetkých riadkov.



Obrázok 4: Orezanie názvu na 128 znakov (Vlastné)

3.2.5 Porovnanie číselníkov 2024 a 2025 (detekcia nových kódov)

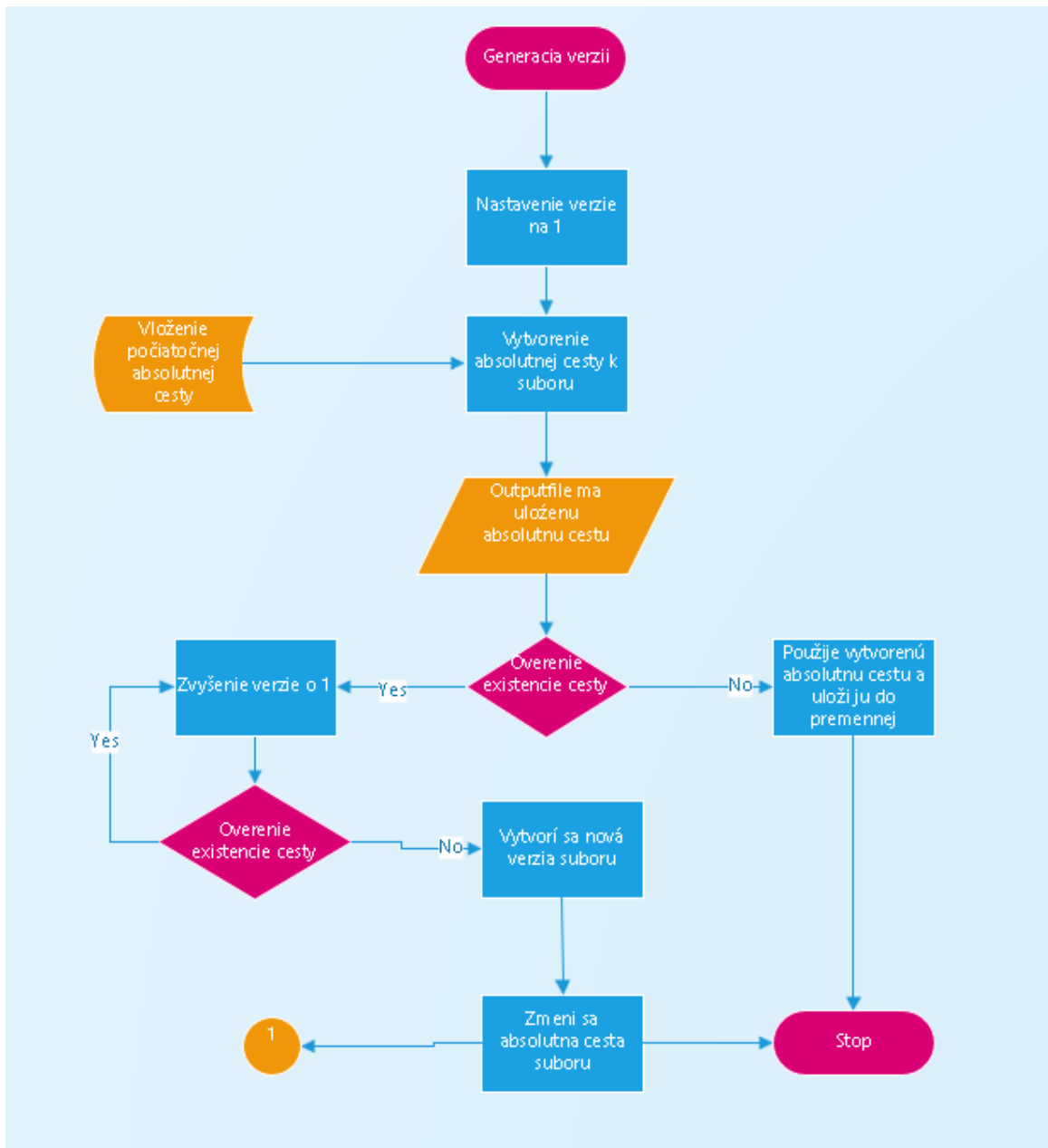
Diagram znázorňuje proces identifikácie nových kódov zdravotníckych výkonov medzi rokmi 2024 a 2025. Na začiatku sa načítajú oba Excel súbory do pamäte a zo súboru pre rok 2024 sa extrahujú existujúce kódy. Následne sa porovnávajú s dátami pre rok 2025. Ak sa nájde nový kód, ktorý neexistoval v zozname z roku 2024, je pridaný do výstupného súboru. Ak kód nie je nový, je riadok ignorovaný. Výsledkom je upravený zoznam s novými kódmi pre rok 2025



Obrázok 5: Detekcia nových kódov (Vlastné)

3.2.6 Zápis výstupných súborov

Diagram opisuje proces generovania verzie výstupného súboru. Na začiatku sa verzia nastaví na hodnotu 1 a vygeneruje sa absolútna cesta k súboru. Systém overí, či cesta už existuje – ak nie, použije sa a uloží do premennej. Ak však cesta existuje, verzia sa zvýši o jedna a kontrola sa opakuje, kým sa nenájde voľná verzia. Po nájdení voľnej verzie sa vytvorí nová cesta a nastaví sa ako aktuálna výstupná.



Obrázok 6: Zápis výstupných súborov (Vlastné)

3.3 Vlastné riešenie

Číselník ZV-DRG výkonov má štruktúru(Tabul'ka č.1):

Tabuľka 1 : Číselník ZV-DRG štruktúra (Vlastné)

Typ ZV-DRG 2024	Kapitola	Kód ZV-DRG 2024	Názov ZV-DRG 2024
1M	1	1-	Diagnostické zdravotné výkony
2M	1	12-	Diagnostické biopsie bez incízie
3M	1	12a-	Biopsie bez incízie centrálného nervového systému
4M	1	12a1-	Biopsie bez incízie intrakraniálnych anatomických štruktúr nervového systému
5M	1	12a10.-	Perkutánná biopsia intrakraniálneho tkaniva s použitím zobrazovacích metód
T	1	12a10.0	Perkutánná biopsia mozgu s použitím zobrazovacích metód
T	1	12a10.1	Perkutánná biopsia bazálnych ganglií s použitím zobrazovacích metód
T	1	12a10.2	Perkutánná biopsia mozgového kmeňa s použitím zobrazovacích metód
T	1	12a10.3	Perkutánná biopsia mozočku s použitím zobrazovacích metód
T	1	12a10.4	Perkutánná biopsia intrakraniálnych častí, hlavových nervov a ganglií s použitím zobrazovacích metód
T	1	12a10.5	Perkutánná biopsia mozgových obalov s použitím zobrazovacích metód
T	1	12a10.x	Perkutánná biopsia ostatného intrakraniálneho tkaniva s použitím zobrazovacích metód

4M	1	12a2-	Biopsie bez incízie miechy, miechových obalov, ganglií a nervov
5M	1	12a20.-	Perkutánná (ihlová) biopsia intraspínálneho tkaniva
T	1	12a20.0	Perkutánná (ihlová) biopsia miechy
T	1	12a20.1	Perkutánná (ihlová) biopsia intraspínálnych častí, miechových nervov a spinálnych ganglií
T	1	12a20.2	Perkutánná (ihlová) biopsia miechových obalov

3.3.1 vybrať len tie riadky, kde hodnoty prvého stĺpca = „T“

Štruktúra vstupného číselníka (Tabuľka č.1)

Číselník ZV-DRG výkonov je štruktúrovaný podľa hierarchie granularity, kde najvyššiu úroveň predstavuje typ „1M“ a najnižšiu úroveň typ „T“. Pre potreby importu do interného informačného systému (IIS) sú relevantné iba kódy s typom „T“, pretože tieto predstavujú konkrétne zdravotné výkony, nie len zoskupené kategórie.

Príprava na ukladanie výsledkov

Ako prvý krok bolo potrebné pripraviť adresáre a názvy výstupných súborov. Skript využíva funkciu `generate_versioned_filename()`, ktorá automaticky vytvára názvy výstupných súborov s pripojeným číslom verzie (napr. `Import_2024_modified_v1.xlsx`), čím sa zabezpečuje, že existujúce dáta nebudú neúmyselne prepísané.

```
output_path_import_modified = generate_versioned_filename(output_dir, base_filename_import + "_modified")
output_path_duplicity = generate_versioned_filename(output_dir, base_filename_duplicity)
```

Obrázok 7: Zápis výstupných súborov (Vlastné)

Načítanie vstupného Excel súboru

Ďalej bol načítaný vstupný Excel súbor. Pre zjednodušenie a urýchlenie spracovania boli načítané len tri relevantné stĺpce: typ výkonu, kód výkonu a názov výkonu.

Týmto krokom sa výrazne znížila pamäťová náročnosť operácie a zvýšila sa rýchlosť spracovania.

```
df = pd.read_excel(file_path, sheet_name=sheet_name, usecols=[type_col, code_col, name_col])
```

Obrázok 8: Načítanie vstupného Excel súboru (Vlastné)

Výber riadkov podľa typu výkonu

Zo všetkých načítaných údajov boli vyfiltrované len tie riadky, kde hodnota v stĺpci typu výkonu bola rovná „T“. Tento filter zabezpečil, že do ďalšieho spracovania sa dostali len zdravotné výkony na najnižšej úrovni podrobnosti.

Týmto spôsobom sa eliminovali záznamy vyšších úrovní hierarchie, ktoré by boli pre import nevhodné.

```
df_filtered = df[df[type_col] == "T"].copy()
```

Obrázok 9: Filtrovanie riadkov (Vlastné)

3.3.2 Kontrola jednoznačnosti kódov a názvov v číselníku ZV-DRG

Jednou z najdôležitejších vlastností každého číselníka je jeho jednoznačnosť. Každý kód v číselníku predstavuje triediacu informáciu o zbieraných dátach, a preto nesmie existovať tzv. duplicita. Skontroloval som, či kódy v stĺpci „Kód ZV-DRG“ sú jednoznačné a neobsahujú opakujúce sa hodnoty. Duplicita kódov by mohla viesť k nesprávnej interpretácii alebo spracovaniu údajov, preto bola táto kontrola nevyhnutnou súčasťou spracovania dát.

Rovnako dôležité bolo overiť aj jednoznačnosť názvov výkonov uvedených v stĺpci „Názov ZV-DRG 2024“. Častou chybou pri tvorbe číselníkov je totiž situácia, keď autor zabezpečí unikátnosť kódov, avšak popisy (názvy) kódov sa opakujú. Pri následnom použití takéhoto číselníka by mohlo dôjsť k skresleniu výsledkov, pretože viaceré kódy by odkazovali na rovnaký názov.

Z tohto dôvodu som vykonal kontrolu duplicit v oboch stĺpcoch – v kódoch aj názvoch – aby bolo zabezpečené správne fungovanie následných analytických a importných procesov.

V našom spracovávanom súbore sme našli duplicitné popisy (Tabuľka č.2).

Tabuľka 2 : Číselník ZV-DRG duplicitné popisy (Vlastné)

Typ ZV-DRG 2024	Kapitola	Kód ZV-DRG 2024	Názov ZV-DRG 2024
T	5	5g607.1	Adenoidektómia endoskopicky
T	K	Kg607.1	Adenoidektómia endoskopicky
T	5	5g607.0	Adenoidektómia otvorene chirurgicky
T	K	Kg607.0	Adenoidektómia otvorene chirurgicky
T	5	5c390	Adheziolýza na priamom očnom svale
T	K	Kc390	Adheziolýza na priamom očnom svale
T	5	5c39a	Adheziolýza na šikmom očnom svale
T	K	Kc39a	Adheziolýza na šikmom očnom svale
T	5	5n52x.2	Anatomická (typická) resekcia pečene, ostatné, laparoskopicky
T	5	5n523.41	Anatomická (typická) resekcia pečene, ostatné, laparoskopicky

```
# Získanie duplicit
duplicate_codes = df_filtered[df_filtered.duplicated(subset=[code_col], keep=False)]
duplicate_names = df_filtered[df_filtered.duplicated(subset=[name_col], keep=False)]
duplicates = pd.concat([duplicate_codes, duplicate_names]).drop_duplicates()
duplicates["Dátum platnosti OD"] = start_date
duplicates["Dátum platnosti DO"] = end_date
num_duplicates = duplicates.shape[0]
```

Obrázok 10: Získanie duplicit (Vlastné)

Načítanie dát (df_filtered):

Celý vstupný Excel súbor je načítaný do pamäte a uložený ako objekt typu DataFrame. Pre efektívnejšie spracovanie sú načítané len vybrané stĺpce, ktoré sú relevantné pre ďalšie operácie.

Identifikácia duplicitných kódov a názvov:

Funkcia `.duplicated(subset=[code_col], keep=False)` prechádza všetky hodnoty v stĺpci s kódmi. Pandas pri tejto operácii:

- prejde hodnoty stĺpca iba raz,
- každú hodnotu si uloží do internej hashovacej tabuľky,
- pri opakovanom výskyte hodnoty zaznamená, že ide o duplicitu.

Rovnaký postup sa aplikuje aj pri kontrole duplícít v názvoch výkonov (`duplicate_names`), tentoraz v stĺpci `name_col`.

Nastavenie parametra `keep=False` zabezpečuje, že ako duplicity budú označené všetky výskyty opakovaných hodnôt.

Spájanie výsledkov:

Pomocou funkcie `pd.concat([...])` sú dva čiastkové výsledky (duplicitné kódy a duplicitné názvy) spojené do jedného objektu `DataFrame`.

Odstránenie opakovaných riadkov:

Funkcia `.drop_duplicates()` odstráni prípadné riadky, ktoré boli zaradené dvakrát – ak boli súčasne duplicitné podľa kódu alebo názvu.

Uloženie informácií o duplícite:

```
duplicate_names_set = set(
    df_filtered[df_filtered.duplicated(subset=['original_name'], keep=False)]['original_name'].values)
```

Obrázok 11: Vytvorenie zoznamu duplícít (Vlastné)

Výsledný zoznam všetkých zistených duplícít je uložený do premennej `duplicates`, a `duplicate_names_set` odkiaľ môže byť následne exportovaný alebo použitý na ďalšiu analýzu.

Tabuľka 3: Porovnanie premenných a ich využitie (Vlastné)

Porovnanie	<code>duplicate_names_set</code>	<code>duplicates</code>
Typ dát	Množina (set)	Tabuľka (DataFrame)
Čo obsahuje	Len hodnoty (názvy) – zoznam duplicitných názvov	Celé riadky (kód, názov, typ, kapitola)
Na čo slúži	Na rýchle overenie duplicity názvu pri spracovaní riadkov	Na export alebo analytickú kontrolu všetkých duplicitných riadkov
Ako sa používa	Používa sa v cykle (<code>process_name(row)</code>)	Používa sa na tvorbu výstupného súboru
Ako vzniká	Vyberú sa duplicitné hodnoty zo stĺpca <code>'original_name'</code>	Spoja sa duplicitné riadky podľa kódov a názvov, odstránia sa duplicitné riadky

Tabuľka porovnáva dve rôzne dátové štruktúry používané pri identifikácii a spracovaní duplicit v číselníkoch. `duplicate_names_set` slúži na rýchle overenie výskytu duplicitných názvov pomocou množiny, zatiaľ čo `duplicates` predstavuje DataFrame obsahujúci celé duplicitné riadky.

Rozdiely sú zrejmé v type uchovávaných dát, v ich použití pri spracovaní a vo vzniku samotných štruktúr. Kým prvá sa využíva v cykloch na kontrolu názvov, druhá slúži na analytické výstupy a export výsledkov.

Problém som na pokyn zadávateľa vyriešil tým, že som pri duplicitných popisoch popisy nahradil zretazením kódu a popisu. Tým som vyriešil podmienku, že importovaný súbor, ktorý bude slúžiť na prenos číselníka do IIS musí mať jednoznačný kód a popis (Tabuľka č.4).

Tabuľka 4: : Číselník ZV-DRG zreťazenie kódu a popisu (Vlastné)

Typ ZV-DRG 2024	Kapitola	Kód ZV-DRG 2024	Názov ZV-DRG 2024
T	5	5g607.1	5g607.1 - Adenoidektómia endoskopicky
T	K	Kg607.1	Kg607.1 - Adenoidektómia endoskopicky
T	5	5g607.0	5g607.0 - Adenoidektómia otvorene chirurgicky
T	K	Kg607.0	Kg607.0 - Adenoidektómia otvorene chirurgicky
T	5	5c390	5c390 - Adheziolýza na priamom očnom svale
T	K	Kc390	Kc390 - Adheziolýza na priamom očnom svale
T	5	5c39a	5c39a - Adheziolýza na šikmom očnom svale
T	K	Kc39a	Kc39a - Adheziolýza na šikmom očnom svale
T	5	5n52x.2	5n52x.2 - Anatomická (typická) resekcia pečene, ostatné, laparoskopicky
T	5	5n523.41	5n523.41 - Anatomická (typická) resekcia pečene, ostatné, laparoskopicky

```

# funkcia na spracovanie názvu a označenie zmien
def process_name(row):
    original = row['original_name']
    code = str(row[code_col])
    is_duplicate = original in duplicate_names_set
    is_too_long = len(original) > 128

    if is_duplicate or is_too_long:
        combined = f"{code} - {original}"
        trimmed = combined[:128]
        return pd.Series([trimmed, "ANO"])
    else:
        return pd.Series([original, "NIE"])

```

Obrázok 12: : Spracovanie popisu (Vlastné)

Táto funkcia s názvom `process_name(row)` slúži na spracovanie názvov v jednom riadku dátovej tabuľky, pričom kontroluje, či názov je duplicitný alebo príliš dlhý. Najprv si z daného riadku načíta pôvodný názov (`original_name`) a číselný kód zo stĺpca `code_col`. Následne overí dve podmienky – či sa názov nachádza v množine duplicitných názvov (tzv. `duplicate_names_set`), a či jeho dĺžka nepresahuje 128 znakov.

Ak je aspoň jedna z týchto podmienok splnená, názov sa spojí s kódom do jedného reťazca v tvare kód – názov, a ak je tento reťazec príliš dlhý, oreže sa na prvých 128 znakov. Výsledkom je dvojica hodnôt: nový upravený názov a označenie "ANO", ktoré signalizuje, že došlo k úprave. Ak názov nebol duplicitný ani príliš dlhý, funkcia jednoducho vráti pôvodný názov spolu s označením "NIE", čím naznačí, že nebolo potrebné nič meniť.

3.3.3 odstrániť znak „,“ z kódu číselníka

Ďalšia podmienka, ktorú vyžaduje IIS do ktorého sa číselník musí vložiť je, že kódy nesmú obsahovať znak „,“.

Odstránil som znak „,“ z kódov. Keďže výstup zapisujem do excelovských súborov(tiež podmienka zadávateľa),tak zvýšenú pozornosť som musel venovať kódom, ktoré obsahujú „,“ (Tabuľka č.5) .

Tabuľka 5 : Číselník ZV-DRG kódy obsahujúce e (Vlastné)

Typ ZV-DRG 2024	Kapitola	Kód ZV-DRG 2024	Názov ZV-DRG 2024
T	3	33011.e	Perfúzne CT vyšetrenie mozgu alebo iných parenchýmových orgánov
T	3	3511e.0	Scintigrafia zdroja krvácania v tráviacom systéme (bez ďalších doplnkových opatrení)
T	3	3511e.2	Scintigrafia zdroja krvácania v tráviacom systéme s doplnkovým snímaním po 24 hodinách
T	3	3511e.3	Scintigrafia zdroja krvácania v tráviacom systéme s doplnkovým SPECT
T	3	3511e.5	Scintigrafia zdroja krvácania v tráviacom systéme s doplnkovým SPECT/CT
T	5	5a112.e	Resekcia intrakraniálneho meningeómu spodiny lebečnej
T	5	5a11e.3	Resekcia schwanómu n. III. intrakraniálne kombinovane
T	5	5a11e.4	Resekcia schwanómu n. IV. intrakraniálne kombinovane

T	5	5a11e.5	Resekcia schwanómu n. V. intrakraniálne kombinovane
T	5	5a11e.6	Resekcia schwanómu n. VI. intrakraniálne kombinovane
T	5	5a11e.7	Resekcia schwanómu n. VII. intrakraniálne kombinovane
T	5	5a11e.8	Resekcia schwanómu n. VIII. intrakraniálne kombinovane
T	5	5a11e.9	Resekcia schwanómu n. IX. intrakraniálne kombinovane

Pri exportovaní upravených údajov do formátu Excel sa objavil problém: Excel automaticky interpretoval niektoré hodnoty ako čísla, čo mohlo viesť k ich zmenám (napr. zobrazenie v exponenciálnom tvare). Tento problém bol vyriešený prevodom všetkých hodnôt na text pred ich exportom, čím sa zabezpečila konzistentnosť kódov aj po otvorení súboru v prostredí Excel.

Tabuľka 6 Číselník ZV-DRG kódy obsahujúce e úprava Excelu (Vlastné)

Typ ZV-DRG 2024	Kapitola	Kód ZV-DRG 2024	Názov ZV-DRG 2024
T	3	33011e	Perfúzne CT vyšetrenie mozgu alebo iných parenchýmových orgánov
T	3	3,51E+03	Scintigrafia zdroja krvácania v tráviacom systéme (bez ďalších doplnkových opatrení)
T	3	3,51E+05	Scintigrafia zdroja krvácania v tráviacom systéme s doplnkovým snímaním po 24 hodinách
T	3	3,51E+06	Scintigrafia zdroja krvácania v tráviacom systéme s doplnkovým SPECT

T	3	3,51E+08	Scintigrafia zdroja krvácania v tráviacom systéme s doplnkovým SPECT/CT
---	---	----------	-------------------------------------------------------------------------

Pri úprave kódov zdravotníckych výkonov bolo potrebné odstrániť bodky, ktoré sa v niektorých prípadoch nachádzali v ich zápise. Na tento účel bol použitý nasledujúci riadok kódu v jazyku Python:

```
# Odstránenie bodiek
df_filtered[code_col] = df_filtered[code_col].astype(str).str.replace(".", "", regex=False)
```

Obrázok 13: Odstránenie bodiek v prostredí Python(Vlastné)

Tento príkaz najprv prevedie hodnoty v danom stĺpci na reťazce pomocou `astype(str)`, čím sa zabezpečí správna manipulácia s údajmi ako s textom. Následne pomocou `str.replace(".", "", regex=False)` sa z každého reťazca odstránia výskyty bodiek. Použitie parametra `regex=False` zabezpečuje, že bodka nie je interpretovaná ako špeciálny znak regulárneho výrazu, ale ako bežný znak.

3.3.4 počet znakov, ktorý má Názov ZV-DRG 2024 nesmie presiahnuť 128

Dĺžka reťazca musí byť nanajvýš 128 znakov. Na pokyn zadávateľa som usekol názov na 128 znakov. Keďže tvorba názvu nepredpokladala takýto zásah, tak názvy kódov sa vo veľa prípadoch líšili na konci reťazca(Tabuľka 7) .

Tabuľka 7: : Číselník ZV-DRG dlhé popisy (Vlastné)

Typ ZV-DRG 2024	Kapitola	Kód ZV-DRG 2024	Názov ZV-DRG 2024
T	5	5b12021	Hemityreoidektómia bez odstránenia prištítnych teliesok s disekciou centrálného kompartmentu a krčnou disekciou laterálneho kompartmentu lymfatických uzlín, miniinvazívna videoasistovaná
T	5	5b12022	Hemityreoidektómia bez odstránenia prištítnych teliesok s disekciou centrálného kompartmentu a krčnou disekciou laterálneho kompartmentu lymfatických uzlín, miniinvazívna radioasistovaná
T	5	5b12023	Hemityreoidektómia bez odstránenia prištítnych teliesok s disekciou centrálného kompartmentu a krčnou disekciou laterálneho kompartmentu lymfatických uzlín, miniinvazívna video a radioasistovaná
T	5	5b12024	Hemityreoidektómia bez odstránenia prištítnych teliesok s disekciou centrálného kompartmentu a krčnou disekciou laterálneho kompartmentu lymfatických uzlín, endoskopicky asistovaná
T	5	5b12025	Hemityreoidektómia bez odstránenia prištítnych teliesok s disekciou centrálného kompartmentu a krčnou disekciou laterálneho kompartmentu lymfatických uzlín, endoskopicky radioasistovaná
T	5	5b12026	Hemityreoidektómia bez odstránenia prištítnych teliesok s disekciou centrálného kompartmentu a krčnou disekciou laterálneho kompartmentu lymfatických uzlín, s použitím elektroneurostimulátora

T	5	5b1202x	Hemityreoidektómia bez odstránenia prištítnych teliesok s disekciou centrálného kompartmentu a krčnou disekciou laterálneho kompartmentu lymfatických uzlín, iný spôsob
---	---	---------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Po urezaní názvu som znovu dostal opakovaný názov – duplicitu, ktorú som vyriešil podobne ako pri kontrole jednoznačnosti kódov.

Urezal som názov na 128 mínus počet znakov kódu a vytvoril názov :

Kód + (128 – počet znakov kódu) znakov názvu a vytvoril som (Tabuľka č.8):

Tabuľka 8: Číselník ZV-DRG dlhé popisy riešenie (Vlastné)

Typ 2024	ZV-DRG	Kapitola	Kód ZV-DRG 2024	Názov ZV-DRG 2024
T	5	5b12021	5b12021	5b12021-Hemityreoidektómia bez odstránenia prištítnych teliesok s disekciou centrálného kompartmentu a krčnou disekciou laterálne
T	5	5b12022	5b12022	5b12022-Hemityreoidektómia bez odstránenia prištítnych teliesok s disekciou centrálného kompartmentu a krčnou disekciou laterálne
T	5	5b12023	5b12023	5b12023-Hemityreoidektómia bez odstránenia prištítnych teliesok s disekciou centrálného kompartmentu a krčnou disekciou laterálne
T	5	5b12024	5b12024	5b12024-Hemityreoidektómia bez odstránenia prištítnych teliesok s disekciou centrálného kompartmentu a krčnou disekciou laterálne
T	5	5b12025	5b12025	5b12025-Hemityreoidektómia bez odstránenia prištítnych teliesok s disekciou centrálného kompartmentu a krčnou disekciou laterálne

T	5	5b12026	5b12026-Hemityreoidektómia bez odstránenia prítitných teliesok s disekciou centrálného kompartmentu a krčnou disekciou laterálne
T	5	5b1202x	5b1202x-Hemityreoidektómia bez odstránenia prítitných teliesok s disekciou centrálného kompartmentu a krčnou disekciou laterálne

Vytvoril som súbor so štruktúrou(Tabul'ka č.9) :

Tabul'ka 9: Výstup z tabulky (Vlastné)

Kód číselníka	Názov číselníka	Dátum platnosti OD	Dátum platnosti DO

Kde kód číselníka a popis som naplnil údajmi , spôsobom, ktorý som vyššie popísal.

Do stĺpca Dátum platnosti OD som zadal dátum : 1.1.2024 , pretože kódy vznikli v roku 2024.

Do stĺpca Dátum platnosti OD som zadal dátum : 31.12.9999, pretože vzniknuté kódy platia v čase, kým sa na pokyn tvorcu neskončí ich platnosť.

Platnosť kódov v čase zabezpečí použitie číselníka pre zber a následný výstup v rokoch ich platnosti.

3.3.5 Aktualizácia číselníka na rok 2025

Číselník Výkonov sa ročne aktualizuje. Súbor ZV_DRG_2025 má rovnakú štruktúru (Tabuľka č.10)

Tabuľka 10: : Číselník ZV_DRG_2025 (Vlastné)

Typ ZV-DRG 2025	Kapitola	Kód ZV-DRG 2025	Názov ZV-DRG 2025
4M	3	3220-	Kontrastné RTG vyšetrenia
5M	3	32208.-	Kontrastné RTG vyšetrenie prsníka
T	3	32208.2	Kontrastná mamografia
5M	3	35136.-	Rádionuklidová cystografia
5M	3	35137.-	Rádionuklidové vyšetrenie plazmatického klírensu
6M	5	5a902.4-	Ventrálny prístup ku krčnej chrbtici
T	5	5a902.41	Ventrálny prístup ku krčnej chrbtici, 1 segment
T	5	5a902.42	Ventrálny prístup ku krčnej chrbtici, 2 segmenty
T	5	5a902.43	Ventrálny prístup ku krčnej chrbtici, 3 a viac segmentov
6M	5	5a902.5-	Ostatné prístupy ku krčnej chrbtici
T	5	5a902.5a	Prístup ku kraniocervikálnemu spojeniu a krčnej chrbtici: krčná chrbtica, laterálne

.....

Kontrolu a prípravu nového súboru na import do IIS som robil v niekoľkých krokoch: aplikoval som pravidlá, ktoré som použil v prvom kroku (Príprava súboru na import), čiže:

- vybrať len tie riadky, kde hodnoty prvého stĺpca = "T"
- skontrolovať jednoznačnosť kódov v stĺpci C- Kód ZV-DRG
- skontrolovať jednoznačnosť kódov v stĺpci D- Názov ZV-DRG 2025
- odstrániť znak „,“ z kódu číselníka
- počet znakov, ktorý má Názov ZV-DRG 2025 nesmie presiahnuť 128

Súbor ZV_DRG_2025 obsahuje všetky platné kódy na rok 2025. Do IIS však potrebujem naimportovať tie kódy, ktoré v roku 2025 pribudli. Musel som súbory porovnať a vybrať len tie kódy, ktoré sú v roku 2025, ale nie sú v roku 2024.

```
import_2024_path = find_latest_version_file(output_dir_2024, "Import_2024_modified")
import_2025_path = find_latest_version_file(output_dir_2025, "Import_2025_modified")
```

Obrázok 14: Nájdenie poslednej verzie upravených import súborov (Vlastné)

Pomocou pomocnej funkcie `find_latest_version_file()` sa v zadaných adresároch vyhľadá najnovšia verzia Excel súborov pre daný rok. Cieľom je, aby sa vždy pracovalo s najaktuálnejšími údajmi bez manuálneho výberu súboru.

```
df_2024 = pd.read_excel(import_2024_path)
df_2025 = pd.read_excel(import_2025_path)
```

Obrázok 15: Načítanie dát z Excelu (Vlastné)

Pomocou `pandas.read_excel()` sa oba súbory načítajú do dátových rámcov (DataFrame). V tomto kroku sa Excel tabuľky prevedú na dátovú štruktúru vhodnú na analýzu.

```
df_2024 = df_2024.rename(columns={"Kód ZV-DRG 2024": "Kód ZV-DRG"})
df_2025 = df_2025.rename(columns={"Kód ZV-DRG 2025": "Kód ZV-DRG"})
```

Obrázok 16: Premenovanie názvy stĺpcov s kódmi na jednotný názov (Vlastné)

V oboch súboroch sa názvy stĺpcov s kódom výkonu premenovávajú na rovnaký názov – Kód ZV-DRG pre porovnanie musia byť názvy stĺpcov identické.

```
codes_2024 = set(df_2024["Kód ZV-DRG"].astype(str))
df_2025_new = df_2025[~df_2025["Kód ZV-DRG"].astype(str).isin(codes_2024)]
```

Obrázok 17: Získanie nových kódov (Vlastné)

`codes_2024` je množina všetkých kódov z roku 2024. Použije sa typ `set`, aby sa zjednodušilo vyhľadávanie a porovnanie.

`df_2025_new` obsahuje len tie riadky z roku 2025, ktorých kódy sa nenachádzajú v `codes_2024`. Takto sa identifikujú nové kódy, ktoré boli pridané len v roku 2025.

```
combined_df = pd.concat([df_2024, df_2025_new], ignore_index=True)
```

Obrázok 18: Spojenie tabuľky (Vlastné)

Pomocou `pandas.concat()` sa vytvorí nový DataFrame, ktorý kombinuje všetky pôvodné údaje z roku 2024 a novinky z roku 2025. Parameter `ignore_index=True` zabezpečí, že výsledný index bude plynulý.

```
combined_df.to_excel(output_final_path, index=False)
```

Obrázok 19: Uloženie do výstupného súboru (Vlastné)

Nakoniec sa výsledný DataFrame uloží do nového Excel súboru. Parameter `index=False` zabezpečí, že sa index pandas tabuľky nezapíše ako nový stĺpec.

Vytvoril som súbor so štruktúrou (Tabuľka 11):

Tabuľka 11: Aktualizácia číselníka na rok 2025 (Vlastné)

Kód číselníka	Názov číselníka	Dátum platnosti OD	Dátum platnosti DO

Kde kód číselníka a popis som naplnil údajmi , spôsobom, ktorý som vyššie popísal.

Do stĺpca Dátum platnosti OD som zadal dátum : 1.1.2025 , pretože kódy vznikli v roku 2025.

Do stĺpca Dátum platnosti OD som zadal dátum : 31.12.9999, pretože vzniknuté kódy platia v čase, kým sa na pokyn tvorcu neskončí ich platnosť.

Pri analýze problému som narazil na problém. Niektoré kódy číselníka, ktoré vznikli v roku 2024 boli zmenené a v roku 2025 boli nahradené novými kódmi(Tabuľka č.12).

Tabuľka 12: Aktualizácia číselníka na rok 2025 (Vlastné)

Typ ZV-DRG	Kód ZV-DRG 2024	Názov ZV-DRG 2024	Kód ZV-DRG 2025	Názov ZV-DRG 2025
T	32203.4	Bronchografia	32209.4	Bronchografia
T	32205	Vazovezikulografia	32204.5	Vazovezikulografia
T	32206.1	Hysterosalpingografia	32204.6	Hysterosalpingografia

T	32206.2	Vaginografia	32204.7	Vaginografia
T	32206.3	Duktografia	32208.0	Duktografia
T	32207	Artrografia	32209.5	Artrografia
T	35136	Rádionuklidová cystografia priama	35136.0	Rádionuklidová cystografia priama
T	5a400.3	Kryoablácia periférneho nervového tkaniva	8f607.2	Kryoablácia periférneho nervového tkaniva
T	5a904.7	Kombinácia pararektal-retroperitoneálneho prístupu	5a904.54	Kombinácia pararektal-retroperitoneálneho prístupu
T	5b421.4	Maximálna miniinvazívna tymektómia [MMIT]	5b421.40	Maximálna miniinvazívna tymektómia [MMIT]
T	5n198	Pyloromyotómia	5n19s.a	Pyloromyotómia
T	5n199	Pyloroplastika	5n19s.b	Pyloroplastika
T	5n19a	Resekcia pyloru s gastroduodenostómiou	5n19s.c	Resekcia pyloru s gastroduodenostómiou

Takáto tvorba nových kódov je chybou. V aktuálnom roku ,v ktorom číselník platí ,je možné použitie na zber dát – čiže vieme použiť kód pri výbere špecifikácie výkonu (keďže je to číselník zdravotných výkonov). Ale pri tvorbe štatistických výstupov takáto tvorba nových kódov neumožňuje porovnávať hodnoty medziročne , keďže jeden rok platí jeden kód a druhý rok sa kód zmení.

Po dohode so zadávateľom som vytvoril súbor – „ analýza kódov „. Je to zoznam kódov, ktorým sa v roku 2025 zmenil kód. Zoznam je štruktúrovaný . K názvu kódu (stl.A) sa do stl.B uvedie kód výkonu platný v roku 2024 a do stl.C kód výkonu platný v roku 2025 (Tabuľka č.13).

Tabuľka 13: Analýza kódov číselníka (Vlastné)

Názov ZV-DRG 2025	Kód ZV-DRG 2024	Kód ZV-DRG 2025
Bronchografia	32203.4	32209.4
Vazovezikulografia	32205	32204.5
Hysterosalpingografia	32206.1	32204.6
Vaginografia	32206.2	32204.7
Duktografia	32206.3	32208.0
Artrografia	32207	32209.5
Rádionuklidová cystografia priama	35136	35136.0
Kryoablácia periférneho nervového tkaniva	5a400.3	8f607.2
Kombinácia pararektal-retroperitoneálneho prístupu	5a904.7	5a904.54
Maximálna miniinvazívna tymektómia [MMIT]	5b421.4	5b421.40
Pyloromyotómia	5n198	5n19s.a
Pyloroplastika	5n199	5n19s.b
Resekcia pyloru s gastroduodenostómiou	5n19a	5n19s.c
Abdominálna rektopexia, laparotomicky	5n433	5n431.0
Abdominálna rektopexia, laparoskopicky	5n434	5n431.2
Rektopexie cez zadnú rektotómiu	5n435	5n431.6

3.4 Význam implementácie skriptu v praxi a odporúčania pre NCZI

Navrhnutý Python skript predstavuje efektívny nástroj, ktorý môže významne zjednodušiť a zrýchliť spracovanie číselníkov ZV-DRG. Jeho nasadenie v prostredí NCZI by mohlo prispieť k zníženiu chybovosti pri importe údajov do informačného systému, eliminácii duplicit a zabezpečeniu jednotnosti názvov a kódov. Tým by sa výrazne zlepšila kvalita dát a zároveň by sa odbremenili zamestnanci od manuálnej kontroly a opakujúcich sa úloh.

Aby bolo možné tento nástroj efektívne využívať v praxi, bude potrebné zaškoliť zamestnancov, ktorí s číselníkmi pracujú. Školenie by sa malo zamerať na správne spúšťanie skriptu, prácu s výstupnými Excel súbormi, interpretáciu výsledkov a pochopenie základných princípov, na ktorých je skript postavený. Zároveň je dôležité zabezpečiť dostupnosť skriptu na centrálnom úložisku spolu s aktuálnym používateľským manuálom, ktorý bude slúžiť ako referenčný dokument pri jeho používaní.

Implementácia tohto riešenia do interných procesov NCZI by predstavovala dôležitý krok smerom k systematizácii správy číselníkov a k ďalšiemu rozvoju dátovo orientovaných a automatizovaných postupov v rámci zdravotníckej informatiky.

Záver

Hlavným cieľom tejto bakalárskej práce bolo analyzovať číselník zdravotníckych výkonov ZV-DRG za roky 2024 a 2025, identifikovať nedostatky súvisiace s importom do interného informačného systému (IIS) NCZI a navrhnúť efektívny spôsob ich automatizovaného spracovania. Sekundárne ciele zahŕňali vytvorenie nástrojov na kontrolu duplicit, odstránenie nekompatibilných znakov, rešpektovanie dĺžkových limitov názvov.

V teoretickej časti práce boli objasnené základné pojmy ako informácia, dáta a databázy, pričom dôraz bol kladený na význam číselníkov v informačných systémoch. Následne boli popísané výhody použitia jazyka Python a knižníc pandas a openpyxl pri spracovaní štruktúrovaných údajov.

Druhá časť práce sa venovala analýze prostredia NCZI – jeho organizačnej štruktúre, technickému zázemiu, informačným systémom a výzvam, ktorým inštitúcia čelí pri spracovaní zdravotníckych údajov. Zvláštna pozornosť bola venovaná praktickým problémom číselníka ZV-DRG, ako sú duplicity kódov a názvov, prekročenie maximálnej dĺžky názvu či výskyt znakov nežiaducich pri importe.

Vo vlastnej časti práce bol navrhnutý a implementovaný skript v jazyku Python, ktorý automatizoval filtrovanie relevantných záznamov, kontrolu jednoznačnosti, orezanie názvov, detekciu nových kódov a odstraňovanie znakov nevhodných pre databázové systémy. Výstupné súbory boli automaticky verzované a pripravené na import do IIS.

Na základe praktickej implementácie môžeme konštatovať, že navrhnutý nástroj výrazne zefektívnil spracovanie číselníkov a minimalizoval riziko chýb, ktoré by pri ručnom spracovaní vznikli. Významnou výhodou riešenia je aj jeho opakovateľnosť a možnosť budúcej integrácie do interných systémov NCZI, čím by sa znížila závislosť na manuálnych kontrolách a zvýšila konzistencia údajov.

Hoci hlavné technické ciele boli naplnené, projekt tiež odhalil potrebu systematickejšieho riadenia verzií číselníkov a centrálnej validácie údajov ešte pred ich zverejnením.

Do budúcnosti možno odporučiť vybudovanie webového rozhrania s integrovanými kontrolami a notifikáciami, ktoré by zabezpečilo vyššiu transparentnosť a efektivitu práce s rozsiahlymi číselníkmi.

Projekt možno považovať za úspešný, keďže splnil zadanie, priniesol konkrétne výsledky a vytvoril základ pre ďalšie zlepšenia v oblasti správy zdravotníckych údajov.

Zoznam použitej literatúry

- (1) KOCH, Miloš a Bernard NEUWIRTH. Datové a funkční modelování. Vyd. 4., rozš. Brno: Akademické nakladatelství CERM, 2010. ISBN 978-80-214-4125-5.
- (2) KŘÍŽ, Jiří a Petr DOSTÁL. Databázové systémy. Vyd. 1. Brno: Akademické nakladatelství CERM, 2005. ISBN 80-214-3064-8.
- (3)McKINNEY, Wes. *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython*. 2nd ed. Sebastopol: O'Reilly Media, 2018. ISBN 978-1-491-95766-0.
- (4)OPENSafely. *What are codelists?* [online]. 2025 [cit. 2025-05-04]. Dostupné z: <https://www.opensafely.org/codelists/>
- HOTEK, M. SQL Server 2008, krok za krokem. Brno, Computer Press, 2009, 488 s. ISBN 978-80-251-2466-6.
- LACKO, Ľuboslav. Mistrovství v SQL Server 2012. Brno: Computer Press, 2013. ISBN 978-80- 251-3773-4.
- OPPEL, Andrew J. SQL bez předchozích znalostí: [průvodce pro samouky]. Brno: Computer Press, 2008. ISBN 978-80-251-1707-1.
- PRATT, PHILIP J.; LAST, MARY Z. A Guide to SQL, Cengage Learning, Inc 2014, 360s., ISBN 9781111527273
- STEPHENS, R.K. a R.R. PLEW. Naučte se SQL za 21 dní. Brno, Computer Press 2004, 491 s., ISBN 80-722-6870-8

Zoznam použitých obrázkov

Obrázok 1: Filtrovanie relevantných záznamov (Vlastné)	26
Obrázok 2: Kontrola duplicit v kódoch a názvoch (Vlastné)	27
Obrázok 3:Odstránenie nekompatibilných znakov (Vlastné).....	28
Obrázok 4: Orezanie názvu na 128 znakov (Vlastné).....	29
Obrázok 5: Detekcia nových kódov (Vlastné).....	30
Obrázok 6: Zápis výstupných súborov (Vlastné).....	31
Obrázok 7: Zápis výstupných súborov (Vlastné).....	33
Obrázok 8: Načítanie vstupného Excel súboru (Vlastné).....	34
Obrázok 9: Filtrovanie riadkov (Vlastné).....	34
Obrázok 10: Získanie duplicit (Vlastné)	35
Obrázok 11: Vytvorenie zoznamu duplicit (Vlastné)	36
Obrázok 12: : Spracovanie popisu (Vlastné)	39
Obrázok 13: Odstránenie bodiek v prostredí Python(Vlastné).....	42
Obrázok 14: Nájdenie posledné verzie upravených import súborov(Vlastné)	47
Obrázok 15: Načítanie dát z Excelu (Vlastné).....	47
Obrázok 16: Premenovanie názvy stĺpcov s kódmi na jednotný názov (Vlastné).....	48
Obrázok 17: Získanie nových kodov (Vlastné)	48
Obrázok 18: Spojenie tabuľky(Vlastné).....	48
Obrázok 19: Uloženie do výstupného súboru(Vlastné)	48

Zoznam použitých tabuliek

Tabuľka 1 : Číselník ZV-DRG štruktúra (Vlastné)	32
Tabuľka 2 : Číselník ZV-DRG duplicitné popisy (Vlastné)	35
Tabuľka 3: Porovnanie premenných a ich využitie (Vlastné)	37
Tabuľka 4: : Číselník ZV-DRG zret'azenie kódu a popisu (Vlastné)	38
Tabuľka 5 : Číselník ZV-DRG kódy obsahujúce e (Vlastné)	40
Tabuľka 6 Číselník ZV-DRG kódy obsahujúce e úprava Excelu (Vlastné)	41
Tabuľka 7: : Číselník ZV-DRG dlhé popisy (Vlastné)	43
Tabuľka 8: Číselník ZV-DRG dlhé popisy riešenie (Vlastné)	44
Tabuľka 9: Výstup z tabulky (Vlastné).....	45
Tabuľka 10: : Číselník ZV_DRG_2025 (Vlastné)	46
Tabuľka 11: Aktualizácia číselníka na rok 2025 (Vlastné).....	49
Tabuľka 12: Aktualizácia číselníka na rok 2025 (Vlastné).....	49
Tabuľka 13: Analýza kódov číselníka (Vlastné).....	51

Zoznam skratiek

CT – Počítačová tomografia (Computed Tomography)

DRG – Diagnosis-Related Groups, klasifikačný systém pre zdravotnícke výkony

ICPC-2 – International Classification of Primary Care, druhé vydanie

IIS – Interný informačný systém (používaný v prostredí NCZI)

ISZI – Informačný systém zdravotníckych indikátorov

NCZI – Národné centrum zdravotníckych informácií

NZP – Národný zdravotný portál

openpyxl – Knižnica v jazyku Python na prácu so súbormi Excel (.xlsx)

pandas – Python knižnica na manipuláciu s tabuľkovými dátami

RTG – Röntgen (röntgenové vyšetrenie)

SQL – Structured Query Language, jazyk pre prácu s databázami

ZV-DRG – Zdravotnícky výkon klasifikovaný v systéme DRG