D13 - Záverečná správa o výsledkoch projektu ALTAML

Projekt financovaný Európskou úniou Next GenerationEU prostredníctvom Plánu obnovy a odolnosti SR v rámci projektu č. 09I05-03-V02-00049.



Obsah

Úvod **3**

1 PhD. kandidáti zameraní na problémy súvisiace s týmto projektom **4**

1.1 Vymedzenie ukazovateľa *4*

1.2 Dávid Varga *4*

1.3 Zoltán Szoplák *4*

1.4 Andrej Oriňák *5*

2 Spoločné projekty podporované inými schémami **6**

2.1 Vymedzenie ukazovateľa *6*

2.2 APVV-21-0336 — Analýza súdnych rozhodnutí metódami umelej inteligencie *6*

2.3 VV-MVP-24-0038 — Analýza zodpovednosti za delikty páchané na internete pomocou metód strojového učenia *7*

3 Výskumné publikácie šíriace výsledky projektu **9**

3.1 Vymedzenie ukazovateľa *9*

3.2 Zoznam publikácií *9*

4 Národné a medzinárodné spolupráce v akademickej sfére a so štátnym aj súkromným sektorom **11**

4.1 Vymedzenie ukazovateľa *11*

4.2 Národné spolupráce *11*

4.3 Medzinárodné spolupráce *12*

5 Aktivity zamerané na šírenie výsledkov pre potenciálnych partnerov a zákazníkov **14**

5.1 Vymedzenie ukazovateľa *14*

5.2 Záverečná konferencia ALTAML *14*

5.3 Konferencia „Právo a inovácie — je právna prax pripravená na umelú inteligenciu?“ *14*

5.4 Workshop „Analýza súdnych rozhodnutí“ *15*

5.5 Workshop „Automatizovaná analýza právnych textov“ na konferencii Míľniky práva v stredoeurópskom priestore 2026 ..... *15*

5.6 Workshop na vedeckej konferencii pre doktorandov a mladých vedeckých pracovníkov „Právo v ére globálnej nestability: odolnosť, bezpečnosť a nové výzvy“ *16*

6 Publikované datasety **17**

6.1 Vymedzenie ukazovateľa *17*

6.2 Zoznam výstupov *17*

7 Publikované metodológie **21**

7.1 Vymedzenie ukazovateľa *21*

7.2 Zoznam výstupov 21

8 Publikované metódy 24

8.1 Vymedzenie ukazovateľa 24

8.2 Zoznam výstupov 24

Úvod

Tento dokument rozpisuje plnenie merateľných ukazovateľov projektu ALTAML. Pre každý ukazovateľ uvádza jeho vymedzenie, spôsob, akým bola stanovená cieľová hodnota, požiadavky na preukazovanie plnenia a konkrétne výsledky dosiahnuté v projekte. Slúži ako podklad pre monitorovacie správy ako aj ako prehľad pre riešiteľov projektu.

Súhrnný prehľad plnenia je uvedený v nasledujúcej tabuľke. Podrobnosti ku každému ukazovateľu obsahujú jednotlivé kapitoly.

Očakávané výsledky a vplyv aktivít projektu	Očakávaný počet	Reálne plnenie
PhD. kandidáti zameraní na problémy súvisiace s týmto projektom	**3**	**3**
Spoločné projekty podporované inými schémami	**2**	**2**
Výskumné publikácie šíriace výsledky projektu	**4**	**5**
Národné spolupráce v akademickej sfére a so štátnym aj súkromným sektorom	**3**	3
Medzinárodné spolupráce v akademickej sfére a so štátnym aj súkromným sektorom	**2**	**3**
Aktivity zamerané na šírenie výsledkov pre potenciálnych partnerov/zákazníkov	**4**	**5**
Publikované datasety	**3**	**3**
Publikované metodológie	**2**	**2**
Publikované metódy	**14**	**14**

1 PhD. kandidáti zameraní na problémy súvisiace s týmto projektom

Očakávaný počet	Reálne plnenie	Stav
3	3	Splnené

1.1 Vymedzenie ukazovateľa

Základom pre odhad tohto ukazovateľa boli aktuálne a plánované doktorandské pozície, ktorých témy odrážajú problémy definované v projekte ALTAML. Ukazovateľ sleduje budovanie výskumnej kapacity v oblasti automatizovanej analýzy právnych textov.

1.2 Automatická extrakcia atribútov a analýza súdnych rozhodnutí (Dávid Varga)

Údaj	Hodnota
Téma dizertačnej práce	Automatická extrakcia atribútov a analýza súdnych rozhodnutí
Študent	RNDr. Dávid Varga
Pracovisko	Ústav informatiky, Prírodovedecká fakulta UPJŠ v Košiciach
Školiteľ	prof. RNDr. Stanislav Krajči, PhD.
Rok vypísania témy	2021
Podporná dokumentácia	Dávid_Varga_zadanie_záverečnej_práce_SK_EN

Väzba na projekt: téma priamo napĺňa vecné jadro pracovného balíka KPB2 - extrakciu atribútov a vzťahov zo súdnych rozhodnutí. Doktorand je spoluautorom publikačných výstupov projektu a spoluviedol workshop „Analýza súdnych rozhodnutí“.

1.3 Použitie metód automatickej sumarizácie príbehov pre analýzu súdnych rozhodnutí (Zoltán Szoplák)

Údaj	Hodnota
Téma dizertačnej práce	Použitie metód automatickej sumarizácie príbehov pre analýzu súdnych rozhodnutí
Študent	RNDr. Zoltán Szoplák

Údaj	Hodnota
Pracovisko	Ústav informatiky, Prírodovedecká fakulta UPJŠ v Košiciach
Školiteľ	prof. RNDr. Stanislav Krajči, PhD.
Rok vypísania témy	2020
Podporná dokumentácia	Zoltán_Szoplák_zadanie_záverečnej_práce_SK_EN

Väzba na projekt: automatická sumarizácia je predpokladom pre sprístupnenie obsahu rozsiahlych súdnych rozhodnutí a nadväzuje na úlohy pracovných balíkov KPB2 a KPB4. Doktorand je spoluautorom publikačného výstupu projektu prezentovaného na konferencii JURIX 2025.

1.4 Dátová analýza legislatívy a judikatúry v obchodnom práve a práve informačných technológií (Andrej Oriňák)

Údaj	Hodnota
Téma dizertačnej práce	Dátová analýza legislatívy a judikatúry v obchodnom práve a práve informačných technológií
Študent	JUDr. Andrej Oriňák
Pracovisko	Právnická fakulta UPJŠ v Košiciach
Školiteľ	doc. JUDr. Regina Hučková, PhD.
Rok vypísania témy	2023
Podporná dokumentácia	Andrej_Oriňák_zadanie_záverečnej_práce_SK_EN

Väzba na projekt: téma prepája právnu analýzu s dátovými metódami a zodpovedá právnej vetve projektu, teda interpretácii atribútov extrahovaných z právnych textov v aplikačnej praxi.

2 Spoločné projekty podporované inými schémami

Očakávaný počet	Reálne plnenie	Stav
2	2	Splnené

2.1 Vymedzenie ukazovateľa

Základom pre odhad boli prebiehajúce a plánované návrhy projektov, ktoré sa týkajú problémov riešených v projekte ALTAML. Ukazovateľ preukazuje, že téma projektu je systematicky rozvíjaná aj mimo jeho vlastného financovania.

2.2 APVV-21-0336 — Analýza súdnych rozhodnutí metódami umelej inteligencie

Údaj	Hodnota
Číslo projektu	APVV-21-0336
Schéma	Agentúra na podporu výskumu a vývoja (APVV) — všeobecná výzva VV 2021
Zodpovedná organizácia	Univerzita Pavla Jozefa Šafárika v Košiciach
Doba riešenia	01.07.2022 - 30.06.2026
Výška príspevku	249 959,00 €

Odkaz na zmluvu: Centrálny register projektov

Cieľ: hlavným cieľom projektu je vytvorenie predpokladov pre významnejšie použitie metód strojového učenia v rámci rozhodovacej činnosti súdov. Parciálnymi cieľmi sú identifikácia a extrakcia atribútov a vzťahov súdnych rozhodnutí a vytvorenie pilotného vyhľadávacieho nástroja na vyhľadávanie relevantných rozhodnutí (vrátane metód na extrakciu explicitných a implicitných atribútov, reprezentácie a uloženia extrahovaných atribútov a vzťahov), a analýza atribútov extrahovaných zo súdnych rozhodnutí.

Anotácia: využívanie umelej inteligencie a inovatívnych technológií vo verejnom sektore pri rozhodovacej činnosti súdov je predmetom aktuálneho záujmu Európskej únie deklarovaného vo viacerých oficiálnych dokumentoch. Rozhodovacia činnosť súdu si vyžaduje znalosť právnych princípov a zásad a porozumenie právnemu textu. Výskumné ciele projektu sú preto zamerané na extrakciu znalostí z textov súdnych rozhodnutí súdov Slovenskej republiky a ich automatizovanú analýzu — vo vzťahu k právnym inštitútom a princípom v aplikačnej praxi, ale aj k spôsobu realizácie dokazovania v súdnom konaní. Riešiteľský tím tvoria výskumníci z

oblastí teórie práva, práva informačných a komunikačných technológií, procesnoprávnych disciplín, dátovej analýzy, počítačových vied a spracovania prirodzeného jazyka.

Súvislosť s projektom ALTAML: projekt rieši rovnaký predmet, a to automatizovanú analýzu súdnych rozhodnutí. Tento projekt je priamym metodickým predchodcom projektu ALTAML. Zhoda je najmä v extrakcii atribútov a vzťahov zo súdnych rozhodnutí (pracovný balík KPB2 projektu ALTAML) a vo vyhľadávaní relevantných rozhodnutí (KPB3 a KPB4). ALTAML na tieto výsledky nadväzuje a rozširuje ich o spracovanie legislatívy, o sémantický graf extrahovaných údajov a o produkčné sprístupnenie výsledkov. Prekrýva sa aj riešiteľská základňa oboch projektov.

2.3 VV-MVP-24-0038 — Analýza zodpovednosti za delikty páchané na internete pomocou metód strojového učenia

Údaj	Hodnota
Číslo projektu	VV-MVP-24-0038
Schéma	APVV — verejná výzva na podporu projektov mladých vedeckých pracovníkov 2024
Zodpovedná organizácia	Univerzita Pavla Jozefa Šafárika v Košiciach
Doba riešenia	01.03.2025 - 31.12.2028
Výška príspevku	170 250 €

Odkaz na zmluvu: Centrálny register projektov

Cieľ: hlavným cieľom projektu je analýza zodpovednosti za súkromnoprávne a verejnoprávne delikty páchané v digitálnom prostredí a spôsobu ich sankcionovania na základe preskúmania relevantných právnych zdrojov s návrhom a využitím metód strojového učenia. Parciálne ciele zahŕňajú vymedzenie právneho rámca zodpovednosti a klasifikáciu kategórií deliktov, vytvorenie kódovaného datasetu extrahovaných právnych textov spolu s návrhom automatizovaných metód extrakcie vybraných atribútov, a napokon analýzu zodpovednosti, komparáciu spôsobov sankcionovania a formulovanie návrhov de lege ferenda.

Anotácia: incidencia protiprávneho konania v digitálnom prostredí internetu neustále stúpa. Zverejňovanie nezákonného obsahu či iné protiprávne činnosti na internete majú potenciál naplniť skutkovú podstatu mnohých súkromnoprávnych alebo verejnoprávnych deliktov. Diskutuje sa nielen o zodpovednosti samotných páchateľov, ale aj o možnosti založenia zodpovednosti sprostredkovateľov, ktorých služby sú na protiprávnu činnosť využívané. V rámci projektu sa predpokladá vytvorenie samostatného kódovaného, resp. anotovaného

datasetu právnych textov, najmä súdnych rozhodnutí a rozhodnutí iných štátnych orgánov, ktorých predmetom sú delikty páchané na internete.

Súvislosť s projektom ALTAML: projekt využíva rovnaké metodické jadro, a to dátovú analýzu právnych textov s implementáciou metód strojového učenia. Tento prístup aplikuje na špecifickú vecnú oblasť. Priamo nadväzuje na anotačné aktivity a na tvorbu kódovaných datasetov právnych textov (KPB1 a KPB2 projektu ALTAML) a rozširuje záber analyzovaných právnych textov o rozhodnutia iných štátnych orgánov než súdov.

3 Výskumné publikácie šíriace výsledky projektu

Očakávaný počet	Reálne plnenie	Stav
4	5	Prekročené

3.1 Vymedzenie ukazovateľa

Základom pre odhad bol počet predchádzajúcich vedeckých článkov publikovaných kľúčovými členmi výskumného tímu. Ukazovateľ zahŕňa recenzované publikácie, ktoré šíria výsledky projektu v odbornej komunite. Publikácie sú výstupom D24 — Recenzované vedecké články.

3.2 Zoznam publikácií

Sokol, P., Bachňáková Rózenfeldová, L., Varga, D., Rudohradská, S., & Hučková, R. (2025). **Anomaly detection in legal text — A case study on cybersecurity contracts**. In R. Markovich, L. Di Caro, A. Rapp, & C. Schifanella (Eds.), Legal Knowledge and Information Systems: JURIX 2025 (Frontiers in Artificial Intelligence and Applications, Vol. 416, pp. 350–355). IOS Press.

Odkaz na článok: https://doi.org/10.3233/FAIA251609

Abstrakt: Rastúca dostupnosť právnej dokumentácie vytvára priestor pre uplatnenie empirických výskumných metód, ktoré umožňujú komplexnú analýzu zložitých právnych dát. Systematická obsahová analýza (Systematic Content Analysis) ponúka štruktúrovaný prístup ku skúmaniu právnych textov. Tento príspevok skúma možnosť detekcie anomálií v právnom texte prostredníctvom využitia metód strojového učenia. Prezentované závery sú formulované na základe prípadovej štúdie zameranej na zmluvy uzatvorené medzi orgánmi verejnej moci a súkromnými poskytovateľmi služieb v oblasti kybernetickej bezpečnosti. V tejto súvislosti navrhujeme prístup k detekcii anomálneho právneho textu, pričom poskytujeme kategorizáciu atribútov, ktoré vplývajú na určenie anomálií, a opisujeme ich význam vo zvolenej doméne.

Horvát, Š., Varga, D., Gurský, P., Fedurcová, N., Baran, S., & Szoplák, Z. (2025). **A self-supervised method for legal reference resolution**. In R. Markovich, L. Di Caro, A. Rapp, & C. Schifanella (Eds.), Legal Knowledge and Information Systems: JURIX 2025 (Frontiers in Artificial Intelligence and Applications, Vol. 416, pp. 300–305). IOS Press.

Odkaz na článok: https://doi.org/10.3233/FAIA251601

Abstrakt: Rozlišovanie referencií na právne predpisy v súdnych rozhodnutiach je náročné vtedy, keď skratky zodpovedajú viacerým kandidátskym právnym predpisom. Zároveň je však kľúčové, pretože spoľahlivé rozlíšenie referencií podporuje vyhľadávanie právnych informácií. Predstavujeme samoučiacu (self-supervised) grafovú metódu, ktorá zo súdnych rozhodnutí konštruuje vážený graf spoluvýskytu, v ktorom uzly predstavujú ustanovenia právnych

predpisov a váhy hrán odrážajú frekvenciu ich spoločného výskytu. Nejednoznačné referencie sa rozlišujú meraním vzdialeností od kontextových ustanovení pomocou viaczdrojového Dijkstrovho algoritmu, pričom sa vyberá najbližší kandidát. Experimenty ukazujú, že navrhnutý prístup zvyšuje presnosť v porovnaní s jednoduchými referenčnými riešeniami, poskytuje skóre spoľahlivosti a predstavuje prakticky použiteľnú metódu rozlišovania. Súčasťou štúdie je aj anotovaný dataset určený pre ďalší výskum.

Sokol, P., & Bachňáková Rózenfeldová, L. (2025). **Perspektívy využitia metódy systematickej kontextovej analýzy v právnom výskume**. Forum Iuris Europaeum, 13(2). ISSN 2644-4364.

Odkaz na článok: http://fie.iuridica.truni.sk/wp-content/uploads/2025/12/FIE-2-2025_na-web.pdf

Abstrakt: Predmetom tohto príspevku je predstaviť metódu systematickej kontextovej analýzy a jej jednotlivé fázy a preskúmať možnosti jej využitia v právnom výskume. V príspevku načrtávame výhody a nevýhody spojené s uplatnením tohto prístupu, pričom zohľadňujeme špecifiká slovenského právneho prostredia dané najmä voľnou dostupnosťou rôznych typov právnych textov. Na základe doterajšej výskumnej práce autorov uvedieme konkrétne príklady využitia tejto metódy, ako aj špecifiká, ktoré sú s ňou spojené.

Rudohradská, S. (2025). **Practical application of the Digital Services Act — the case of the Slovak Republic.** EU and Comparative Law Issues and Challenges Series (ECLIC). Príspevok prezentovaný na medzinárodnej vedeckej konferencii „Strong and Secure Europe: Legal and Economic Aspects“.

Odkaz na článok: https://doi.org/10.25234/eclic/38129

Abstrakt: Aktu o digitálnych službách (Digital Services Act) sa v súčasnosti venuje značná pozornosť nielen zo strany odborníkov, ale aj širokej verejnosti. Ako je uvedené v článku 93 DSA, nariadenie sa v plnom rozsahu uplatňuje od 17. februára 2024. Členské štáty boli povinné začleniť jeho ustanovenia do svojich vnútroštátnych právnych poriadkov a začať s jeho praktickou implementáciou. Predložený príspevok sa venuje otázkam spojeným s implementáciou Aktu o digitálnych službách do právneho poriadku Slovenskej republiky, ktorá DSA prevzala prostredníctvom novely zákona č. 264/2022 Z. z. o mediálnych službách. V podmienkach Slovenskej republiky sme v súčasnosti svedkami prvých rozhodnutí Rady pre mediálne služby, ktorými bola uložená povinnosť odstrániť nezákonný obsah. Cieľom príspevku je preskúmať proces implementácie a aplikácie Aktu o digitálnych službách a analyzovať relevantné rozhodnutia Rady pre mediálne služby so zameraním na komparáciu koncepcie nezákonného obsahu podľa vnútroštátneho slovenského práva a práva Európskej únie.

Szoplák, Z., Varga, D., Gurský, P., Horvát, Š., & Krajči, S. (2025). **Enhancing document retrieval in large corpora: A keyphrase and reference-based approach.** In K. Satoh & K. Nitta (Eds.), *Local Proceedings of the 19th International Workshop on Juris-Informatics (JURISIN 2025) in association with the 17th JSAI International Symposia on AI (JSAI-isAI 2025)*, May 26–27, 2025.

Odkaz na článok:
https://jurisinformaticscenter.github.io/jurisin2025/jurisin2025localproceedings.pdf

Abstrakt: Vyhľadávanie dokumentov, najmä v rozsiahlych korpusoch, zostáva napriek nedávnemu pokroku v oblasti spracovania prirodzeného jazyka (NLP) a veľkých jazykových modelov (LLM) náročnou úlohou. Kontextové okná týchto modelov často nie sú dostatočne veľké na to, aby do nich bolo možné umiestniť veľký počet dokumentov. Na riešenie tohto problému využívajú mnohé úlohy NLP — napríklad vyhľadávanie dokumentov pre systémy RAG — dvojkrokový postup: najprv sa odfiltrujú najrelevantnejšie dokumenty a tie sa následne použijú ako vstup pre jazykový model, ktorý ich zoradí. Čím je však korpus rozsiahlejší, tým väčšia je potreba robustnej metódy vyhľadávania. Autori navrhujú riešenie, ktoré štandardnú metódu vyhľadávania založenú na podobnosti vektorových reprezentácií rozširuje o skóre vychádzajúce zo zhody extrahovaných kľúčových fráz a ďalej o rozšírenie využívajúce referencie v právnych dokumentoch. Prvý prístup dosiahol strednú priemernú presnosť (mAP) 80,25 %, druhý tento výsledok zlepšil na 92,85 %.

4 Národné a medzinárodné spolupráce v akademickej sfére a so štátnym aj súkromným sektorom

Typ spolupráce	Očakávaný počet	Reálne plnenie	Stav
Národné spolupráce	3	3	Splnené
Medzinárodné spolupráce	2	3	Prekročené

4.1 Vymedzenie ukazovateľa

Základom pre odhad boli už existujúce spolupráce a potenciál pre budúce spolupráce s entitami, ktoré majú prirodzený záujem o výsledky projektu.

Ku každej z uvedených organizácií je podpísané Potvrdenie o záujme spolupracovať.

4.2 Národné spolupráce

Organizácia	Sektor
Ministerstvo spravodlivosti Slovenskej republiky	Štátny sektor

Organizácia	Sektor
Ústav štátu a práva SAV, v. v. i.	Akademická sféra
Slovenská advokátska komora	Profesijná komora

Ministerstvo spravodlivosti Slovenskej republiky je ústredným orgánom štátnej správy pre súdnictvo a väzenstvo a správcom informačných systémov justície vrátane zdrojov súdnych rozhodnutí. Ako gestor digitalizácie justície je prirodzeným používateľom výsledkov projektu pri sprístupňovaní a analýze judikatúry.

Ústav štátu a práva SAV, v. v. i. je vedecké pracovisko Slovenskej akadémie vied zamerané na základný výskum v oblasti teórie práva, ústavného práva a právnych dejín. Jeho výskumné zameranie je komplementárne k právnej vetve projektu, najmä pri interpretácii atribútov extrahovaných z právnych textov.

Slovenská advokátska komora je samosprávnou stavovskou organizáciou zriadenou zákonom o advokácii, ktorá združuje všetkých advokátov Slovenskej republiky, vedie zoznam advokátov a zabezpečuje ich vzdelávanie a disciplinárnu zodpovednosť. Advokáti predstavujú najpočetnejšiu cieľovú skupinu koncových používateľov právnych informačných systémov, čo potvrdil aj prieskum realizovaný v rámci výstupu D01, a komora je preto prirodzeným partnerom pri overovaní praktickej použiteľnosti výsledkov projektu a pri ich šírení do advokátskej praxe.

4.3 Medzinárodné spolupráce

Organizácia	Krajina
Univerzita Palackého v Olomouci, Právnická fakulta	Česká republika
Ústav státu a práva AV ČR, v. v. i.	Česká republika
Tallinn University of Technology (TalTech)	Estónsko

Univerzita Palackého v Olomouci, Právnická fakulta je jednou z popredných českých právnických fakúlt s dlhodobým zameraním na právo informačných technológií a na využitie moderných technológií v právnej praxi. Jej zástupca vystúpil na konferencii „Právo a inovácie“ organizovanej v rámci projektu.

Ústav státu a práva AV ČR, v. v. i. je vedecké pracovisko Akadémie vied Českej republiky zamerané na teoretický a interdisciplinárny právny výskum vrátane práva a technológií. Spolupráca umožňuje medzinárodnú komparáciu prístupov k automatizovanej analýze judikatúry.

Tallinn University je estónska verejná univerzita, ktorá pôsobí v prostredí jednej z najdigitalizovanejších verejných správ v Európe a venuje sa digitálnej transformácii a interdisciplinárnemu výskumu. Spolupráca prináša do projektu skúsenosti z rozvinutého e-Governmentu a digitalizovanej justície.

5 Aktivity zamerané na šírenie výsledkov pre potenciálnych partnerov a zákazníkov

Očakávaný počet	Reálne plnenie	Stav
4	5	Prekročené

5.1 Vymedzenie ukazovateľa

Základom pre odhad boli plánované aktivity na šírenie výsledkov v rámci projektu, teda konferencie a workshopy. Aktivity napĺňajú výstupy pracovného balíka KPB5, a to D30 - Konferencia a D31 - Workshop.

5.2 Záverečná konferencia ALTAML

Údaj	Hodnota
Dátum	24. jún 2026
Miesto	Amfiteáter, 2. poschodie starej budovy, Právnická fakulta Univerzity Komenského v Bratislave
Typ	Konferencia

Odkaz: https://altaml.upjs.sk/news/konferencia-altaml-otvorila-diskusiu-o-zodpovednom-vyuzivani-umelej-inteligencie-v-sudnictve/

Anotácia: Konferencia bola venovaná výsledkom projektu a odbornej diskusii o možnostiach, rizikách a podmienkach zodpovedného využívania umelej inteligencie v súdnictve. Podujatie prepojilo pohľad verejnej správy, súdnej praxe a akademického prostredia. Kľúčovou časťou bol okrúhly stôl „AI na súdoch: podpora rozhodovania, vzdelávanie a hranice ľudskej kontroly“, ktorý sa zameral na využitie umelej inteligencie pri právnom výskume, práci s judikatúrou, sumarizácii spisov a vzdelávaní sudcov a súdnych úradníkov. Diskusia sa venovala aj rizikám generatívnej umelej inteligencie — netransparentnosti, nepresnostiam, halucináciám a automatizačnému skresleniu — a potrebe zachovania zmysluplnej ľudskej kontroly.

5.3 Konferencia „Právo a inovácie — je právna prax pripravená na umelú inteligenciu?“

Údaj	Hodnota
Dátum	14. november 2025

Údaj	Hodnota
Miesto	Právnická fakulta UPJŠ, Kováčska 26, Košice
Typ	Konferencia

Odkaz: https://altaml.upjs.sk/news/je-pravna-prax-pripravena-na-ai-konferencia-pravo-a-inovacie-otvorila-klucove-temy/

Anotácia: Konferenciu zorganizovala Katedra obchodného práva a hospodárskeho práva v spolupráci s Ústavom informatiky Prírodovedeckej fakulty. Podujatie spojilo odborníkov z akademického prostredia, advokácie, súdnictva a technologickej sféry. V expertnom fóre vystúpili zástupcovia Ministerstva spravodlivosti SR, Právnickej fakulty UK, Slovenskej advokátskej komory a Univerzity Palackého v Olomouci. Panelové diskusie sa venovali uplatneniu umelej inteligencie v každodennej právnej práci, automatizácii právnych procesov, budúcnosti právnických povolaní a vzdelávaniu. Súčasťou programu bola prezentácia priebežných výsledkov projektu ALTAML a predstavenie nástrojov partnera Wolters Kluwer.

5.4 Workshop „Analýza súdnych rozhodnutí“

Údaj	Hodnota
Dátum	14. máj 2025
Miesto	Prírodovedecká fakulta UPJŠ, Jesenná 5, Košice
Typ	Workshop

Odkaz: https://altaml.upjs.sk/news/workshop-analyza-sudnych-rozhodnuti-pomocou-dat-a-strojoveho-ucenia/

Anotácia: Workshop viedli členovia riešiteľského kolektívu – RNDr. Dávid Varga a JUDr. Simona Rudohradská, PhD. Účastníci sa oboznámili s využitím dátovej analytiky a metód strojového učenia pri skúmaní súdnych rozhodnutí s dôrazom na grafové štruktúry, ktoré umožňujú reprezentovať jednotlivé rozhodnutia ako uzly a vizualizovať ich vzájomné vzťahy. Prakticky si vyskúšali reprezentáciu rozhodnutí ako grafu, vizualizáciu vzťahov medzi rozhodnutiami a identifikáciu anomálnych rozhodnutí metódou outlier detection. Záver patril ukážke metódy COIN (Contextual Outlier Interpretation). Podujatie bolo určené študentom, výskumníkom a odborníkom so záujmom o právo, dátovú analytiku a spracovanie prirodzeného jazyka.

5.5 Workshop „Automatizovaná analýza právnych textov“ na konferencii Míľniky práva v stredoeurópskom priestore 2026

Údaj	Hodnota
Dátum	27. marec 2026
Miesto	Účelové zariadenie Kancelárie Národnej rady SR Častá – Papiernička
Typ	Workshop

Odkaz: https://altaml.upjs.sk/news/workshop-automated-analysis-of-legal-texts/

Anotácia: Workshop viedol JUDr. Zoltán Gyurász, PhD. Venoval sa špecifikám právnych textov, ktoré sú síce formálne presné, no z pohľadu automatizovaného spracovania predstavujú komplexné štruktúry silne závislé od kontextu. Pozornosť bola zameraná na posun od tradičných prístupov založených na kľúčových slovách k pokročilým metódam počítačového porozumenia — spracovaniu prirodzeného jazyka, extrakcii informácií a metódam strojového učenia. Súčasťou boli ukážky riešení inšpirovaných výskumnými aktivitami projektu ALTAML a diskusia o potrebe kritického prístupu k umelej inteligencii najmä vo vzťahu k zaujatosti a transparentnosti.

5.6 Workshop na vedeckej konferencii pre doktorandov a mladých vedeckých pracovníkov „Právo v ére globálnej nestability: odolnosť, bezpečnosť a nové výzvy“

Údaj	Hodnota
Dátum	30. apríl 2025
Miesto	Právnická fakulta UPJŠ, Kováčska 26, Košice
Typ	Workshop

Odkaz: https://altaml.upjs.sk/news/altaml-workshop-na-konferencii-pre-doktorandov/

Anotácia: Súčasťou konferencie bol odborný workshop venovaný využitiu strojového učenia pri analýze súdnych rozhodnutí, ktorý viedla Simona Rudohradská z Katedry obchodného práva a hospodárskeho práva. Workshop ukázal, že prepájanie práva, dátovej analytiky a umelej inteligencie už nie je víziou budúcnosti, ale realitou, ktorá postupne mení spôsob práce s právnymi textami. Konferencia aj workshop sú výstupom riešenia projektu ALTAML.

6 Publikované datasety

Očakávaný počet	Reálne plnenie	Stav
3	3	Splnené

6.1 Vymedzenie ukazovateľa

Základom pre odhad bol počet datasetov, ktoré je potrebné pripraviť na dosiahnutie cieľov projektu. Do ukazovateľa spadajú výstupy typu „súbor údajov“ určené na zverejnenie.

6.2 Zoznam výstupov

Číslo	Výstup	Pracovný balík	Prístup
D04	Vzorové údaje	KPB1.4	Verejný
D08	Databáza slovenských a anglických právnych fráz	KPB2.4	Verejný
D14	Semantický graf extrahovaných údajov	KPB3.1	Verejný

Očakávaná hodnota troch publikovaných datasetov bola dosiahnutá v plnom rozsahu. Všetky tri datasety sú určené na verejné sprístupnenie v otvorených úložiskách v súlade so zásadami otvorenej vedy a s princípmi FAIR.

6.3 D04 - Vzorové údaje

Údaj	Hodnota
Číslo výstupu	D04
Pracovný balík / úloha	KPB1, úloha KPB1.4
Zodpovedný partner	Wolters Kluwer SR s. r. o.
Typ	súbor údajov (dataset)
Prístup a šírenie	verejný — publikovanie v otvorených úložiskách
Dokument s podrobným popisom	ALTAML-D04-KPB1.4-Vzorove-udaje (SK)
Dostupnosť	publikované na webovom sídle projektu https://altaml.upjs.sk/

Zhrnutie: Výstup predstavuje demonštračnú sadu súdnych rozhodnutí sprístupnenú súbežne v troch formátoch — pôvodné PDF, transformované XML a výsledné HTML — spolu s kompletnou sadou metadát (zdroj, URL, interné ID, spisová značka, dátum, typ a názov rozhodnutia, sumár, vydávajúci súd a názvy jednotlivých súborov). Dataset pokrýva sedem zdrojov rozhodovacej činnosti: portál Ministerstva spravodlivosti SR (justice.gov.sk), webové sídla Najvyššieho súdu SR, Najvyššieho správneho súdu SR a Ústavného súdu SR, ako aj tri zbierky rozhodnutí a stanovísk týchto súdov. Vzhľadom na rozsah spracovaných zdrojov je na demonštračné účely zverejnená len reprezentatívna časť spracovaných dokumentov. Neoddeliteľnou súčasťou výstupu sú dve sprievodné dokumentácie dátových balíkov: časť I popisuje 17 dátových balíkov pre domény autorské právo a ochranné známky/nekalá súťaž, časť II opisuje 14 dátových balíkov pre doménu digitálny dôkaz v trestnom konaní. Ku každému z 31 balíkov je priradená výsledková sada vo formáte .xlsx (datovy-balik-1.xlsx až datovy-balik-31.xlsx) s hárkom „altaml“ obsahujúcim všetky dokumenty a vo väčšine prípadov aj s hárkom „po_redukcii“ so zúženým výberom podľa analýzy spisovej značky (registre „Ca“ a „Co“ pre autorské právo, „CbPV“ a „CoPV“ pre ochranné známky, „T“, „To“, „Tdo“, „Tk“, „Nt“ a „ÚS“ pre trestnú agendu).

Väzba na ciele projektu: Výstup napĺňa špecifický cieľ 1 projektu (identifikácia vhodnej metódy spracovania právneho textu vo vybraných prípadových štúdiách) a zároveň tvorí vstupnú dátovú základňu pre všetky nadväzujúce pracovné balíky KPB2 až KPB5.

6.4 D08 - Databáza slovenských a anglických právnych fráz

Údaj	Hodnota
Číslo výstupu	D08
Pracovný balík / úloha	KPB2, úloha KPB2.4
Zodpovedný partner	Univerzita Pavla Jozefa Šafárika v Košiciach
Typ	súbor údajov (dataset)
Prístup a šírenie	verejný — publikovanie v otvorených úložiskách
Dokument s podrobným popisom	ALTAML-D08-KPB2.4-Databaza-slovenskych-a-anglickych-pravnych-fraz (SK)
Dostupnosť	dataset a sprievodná dokumentácia v repozitári GitLab UPJŠ

Zhrnutie: Výstupom je bilingválna databáza právnych pojmov a ich definícií publikovaná súbežne vo formátoch .csv, .xlsx a .json. Pojmy boli zozbierané z verejne dostupných online zdrojov (Slov-Lex, Najpravo.sk) a doplnené o vybrané nadpisy paragrafov Zbierky zákonov Slovenskej republiky; keďže nie každý nadpis paragrafu je právnym pojmom, nadpisy boli filtrované kombináciou štatistických a sémantických metrík a promptovaním veľkých jazykových modelov. Súbory pravne_pojmy_sk_en.csv a pravne_pojmy_sk_en.xlsx obsahujú

dvojjazyčné zoznamy pojmov (stĺpce SK a EN), zatiaľ čo súbor legal_terms_sk_en.json obsahuje rozšírenú štruktúru s poľami _id (slovenský pojem), definitions (zoznam definícií so zdrojom, pri zdroji „zz“ aj s identifikáciou zákona a nadpisu v jeho štruktúre) a english (anglický ekvivalent a jeho definícia). Tam, kde bol text paragrafu vyhodnotený ako definičný, bola definícia extrahovaná priamo z jeho znenia. Anglické ekvivalenty a definície boli vytvorené prekladom a sumarizáciou slovenských definícií pomocou veľkého jazykového modelu LLaMA 3.3, pričom ich primárnym využitím nie je presnosť definície pre človeka, ale tvorba vektorových reprezentácií pojmov, kde je rozhodujúci kontext a tematická konzistentnosť.

Väzba na ciele projektu: Výstup napĺňa špecifický cieľ 2 projektu (extrakcia explicitných a implicitných atribútov z právneho textu). Databáza tvorí uzavretú množinu kandidátnych pojmov pre výstup D09 — Metóda na priraďovanie kľúčových fráz, čím zabezpečuje terminologickú presnosť a konzistentnosť extrakcie naprieč celým korpusom.

6.5 D14 — Sémantický graf extrahovaných údajov

Údaj	Hodnota
Číslo výstupu	D14
Pracovný balík / úloha	KPB3, úloha KPB3.1
Zodpovedný partner	Univerzita Pavla Jozefa Šafárika v Košiciach
Typ	súbor údajov (dataset)
Prístup a šírenie	verejný — publikovanie v otvorených úložiskách
Dokument s podrobným popisom	ALTAML-D14-KPB3.1-Semanticky-graf (SK)
Dostupnosť	výsledný RDF graf v súbore model.rdf; ontologické prefixy https://court.ics.upjs.sk/ontology/ a https://court.ics.upjs.sk/resource/

Zhrnutie: Výstupom je jednotný RDF graf obsahujúci metadáta rozhodnutí Súdneho dvora Európskej únie aj slovenských súdov v spoločnom sémantickom formáte, vrátane RDFS inferencie, publikovaný v súbore model.rdf. Návrh vychádza z analýzy ontologickej vrstvy Úradu pre publikácie EÚ, ktorú tvoria Common Data Model (CDM) pre štruktúru a väzby dokumentov, SKOS/EuroVoc pre tematické zaradenie a Dublin Core Terms pre bibliografické atribúty. Keďže Ministerstvo spravodlivosti SR zverejňuje nielen judikáty, ale aj bežné rozhodnutia prvého a druhého stupňa, ktoré CDM nepozná a ktoré majú vlastné atribúty (okres a kraj súdu, povaha a forma rozhodnutia, oblasť a podoblasť právnej úpravy), zaviedol riešiteľský tím novú nadradenú triedu upjs_ontology:court_decision v hierarchii cdm:case-law_work ⊑ upjs_ontology:court_decision ⊑ cdm:Work ⊑ owl:Thing. Rozšírenie sa realizovalo výhradne „smerom nahor“ prostredníctvom subClassOf a subPropertyOf, čím zostáva integrita pôvodného modelu CDM nedotknutá a každý európsky judikát sa automaticky stáva

platnou inštanciou novej triedy. Boli navrhnuté štyri nové triedy (decision_nature, decision_form, legal_area, legal_sub_area) a trinásť nových vlastností pokrývajúcich ECLI, spisovú značku, identifikačné číslo spisu, väzby na sudcu, súd a štát pôvodu, hierarchiu okres → kraj a typ súdu.

Väzba na ciele projektu: Výstup napĺňa špecifický cieľ 3 projektu (reprezentácia extrahovaných údajov z právneho textu). Preukazuje, že dátový model používaný inštitúciami EÚ je dostatočne univerzálny aj pre slovenskú judikatúru bez potreby ďalšej konverzie, čím vytvára predpoklad pre vznik jednotnej, interoperabilnej databázy rozhodnutí.

7 Publikované metodológie

Očakávaný počet	Reálne plnenie	Stav
2	2	Splnené

7.1 Vymedzenie ukazovateľa

Základom pre odhad bol počet metodológií, ktoré je potrebné pripraviť na dosiahnutie cieľov projektu. Na rozdiel od jednotlivých metód a algoritmov ide o ucelené metodické rámce, ktoré určujú, ako sa k danej úlohe pristupuje.

7.2 Zoznam výstupov

Číslo	Výstup	Pracovný balík	Typ	Prístup
D01	Prieskum u koncových používateľov	KPB1	prieskum	Verejný
D15	Návrh úložiska	KPB3	návrh	Verejný

Očakávaná hodnota dvoch publikovaných metodológií bola dosiahnutá v plnom rozsahu. Metodický rámec systematickej kontextovej analýzy práva, ktorý bol v projekte overený a vyhodnotený vo výstupe D23 — Hodnotiaca správa, bol navyše publikovaný aj vo forme recenzovaného vedeckého článku (kapitola 3).

7.3 D01 - Prieskum u koncových používateľov

Údaj	Hodnota
Číslo výstupu	D01
Pracovný balík / úloha	KPB1, úloha KPB1.1
Zodpovedný partner	Wolters Kluwer SR s. r. o.
Typ	prieskum / metodický rámec
Prístup a šírenie	verejný — publikovanie na webovom sídle projektu
Dokument s podrobným popisom	ALTAML-D01-KPB1.1-Prieskum-u-koncovych-pouzivatelov (SK), ALTAML-D01-KPB1.1-End-users-survey (EN)
Dostupnosť	https://altaml.upjs.sk/wp-content/uploads/sites/29/2025/07/Uzivatelsky-

Údaj	Hodnota
	prieskum_altaml_WK_2025_04_04.pdf a https://altaml.upjs.sk/wp-content/uploads/sites/29/2025/07/ALTAML_Users_survey_en.pdf

Zhrnutie: Cieľom prieskumu bolo identifikovať potreby, zvyklosti a očakávania subjektov, ktoré pravidelne pracujú s právnymi textami, a zistiť, ako koncoví používatelia vnímajú súčasné právne informačné nástroje, čo im v nich chýba a aké funkcionality by zvýšili efektivitu ich práce. Prieskumu sa zúčastnilo 102 právnych profesionálov z rôznych oblastí práva a regiónov Slovenskej republiky — advokáti a právnici v súkromnej sfére, in-house právnici vo firmách a zamestnanci verejnej správy a justície. Zber údajov kombinoval dotazník so štruktúrovanými a otvorenými otázkami s priamymi hĺbkovými rozhovormi s vybranými respondentmi; dotazníky boli anonymné a údaje boli spracované v súlade s nariadením (EÚ) 2016/679. Po ukončení zberu prebehlo posúdenie relevantnosti odpovedí, pričom boli vyradené neúplné odpovede, prázdne formuláre a odpovede bez výpovednej hodnoty. Prieskum priniesol tri kľúčové zistenia: potrebu rýchleho, presného a dôveryhodného vyhľadávania, ktorého výsledok je ihneď použiteľný bez ďalšieho overovania; dopyt po stručných zhrnutiach súdnych rozhodnutí a po extrakcii kľúčových slov nad rozhodnutiami; a opakovanú požiadavku na nástroj na predikciu výsledku právneho sporu. Súčasťou výstupu sú dve prezentácie — vlastný prieskum medzi 102 respondentmi a medzinárodný prieskum 2024 Wolters Kluwer Future Ready Lawyer Survey realizovaný medzi 712 právnikmi v USA a deviatich európskych krajinách, ktorý tvorí širší referenčný rámec.

Väzba na ciele projektu: Výstup napĺňa špecifický cieľ 1 projektu a metodicky usmerňuje všetky nasledujúce pracovné balíky. Tri identifikované požiadavky boli priamo premietnuté do výstupov projektu: požiadavka na vyhľadávanie do D16 a D17, požiadavka na zhrnutia a kľúčové slová do D06, D09 a D10, a požiadavka na predikciu výsledku sporu do D21.

7.4 D15 - Návrh úložiska

Údaj	Hodnota
Číslo výstupu	D15
Pracovný balík / úloha	KPB3, úloha KPB3.2
Zodpovedný partner	Univerzita Pavla Jozefa Šafárika v Košiciach
Typ	návrh architektúry / metodický rámec
Prístup a šírenie	verejný — publikovanie v otvorených úložiskách
Dokument s podrobným popisom	ALTAML-D15-KPB3.2-Navrh-repozitara (SK)
Dostupnosť	dokumentácia architektúry; komponenty dostupné v repozitároch GitLab UPJŠ

Zhrnutie: Výstup definuje modulárnu dátovú architektúru systému pre automatizované spracovanie, obohacovanie a sprístupňovanie súdnych rozhodnutí. Základom je komponent ALTAML sťahovač rozhodnutí, ktorý zabezpečuje získavanie rozhodnutí z externých zdrojov, ich konverziu, verziovanie a postupné aplikovanie nezávislých extrakčných metód. Architektúra kombinuje tri typy úložísk podľa fázy spracovania: relačná databáza MySQL plní riadiacu a referenčnú funkciu s centrálnou tabuľkou raw_decision, číselníkmi súdov a uloženými JSON schémami externých API, čím sa znižuje zaťaženie serverov Ministerstva spravodlivosti; dokumentová databáza MongoDB uchováva finálne obohatené rozhodnutia vrátane histórie verzií extrakčných metód, čo umožňuje návrat k staršej verzii pri chybnej metóde aj spustenie novej metódy nad už spracovanými rozhodnutiami bez opakovania celého procesu; a Elasticsearch tvorí finálnu vyhľadávaciu vrstvu s vlastným analyzérom postaveným na slovenskom Hunspell slovníku. Návrh ďalej popisuje štyri samostatné služby a ich repozitáre — Tvaroslovník API pre lematizáciu a gramatické kategórie, Referex pre extrakciu odkazov s piatimi podpornými MongoDB kolekciami, Keyphrase Extractor pre identifikáciu kľúčových pojmov a Annotate ako vyhľadávacie a anotačné rozhranie. Dokument zároveň transparentne zaznamenáva metodickú zmenu oproti pôvodnému zámeru: návrh vlastného dopytovacieho jazyka a samostatnej middleware vrstvy bol nahradený využitím natívneho dopytovacieho mechanizmu Elasticsearch, ktorý sa ukázal ako dostatočný a navyše umožňuje kombinovať atribútové, fulltextové a sémantické vyhľadávanie.

Väzba na ciele projektu: Výstup napĺňa špecifický cieľ 3 projektu (reprezentácia extrahovaných údajov a návrh vhodnej metódy vyhľadávania). Architektúra je priamym predpokladom pre výstupy D16 a D17 a zároveň zabezpečuje škálovateľnosť, opakovateľnosť spracovania a udržateľnosť výsledkov po ukončení projektu.

8 Publikované metódy

Očakávaný počet	Reálne plnenie	Stav
14	14	Splnené

8.1 Vymedzenie ukazovateľa

Základom pre odhad bol počet metód a algoritmov, ktoré je potrebné pripraviť na dosiahnutie cieľov projektu. Do ukazovateľa spadajú výstupy typu „metóda“, „algoritmus“ a „model“ naprieč pracovnými balíkmi KPB1 až KPB4. Všetky sú určené na zverejnenie.

8.2 Zoznam výstupov

Číslo	Výstup	WP	Typ
D02	Metóda spracovania právnych textov	KPB1	metóda
D03	Metódy pre kontinuálne sťahovanie	KPB1	metóda
D05	Metóda na extrakciu referencií	KPB2	metóda
D06	Metóda na extrakciu rozsudkov	KPB2	metóda
D07	Metóda na extrakciu procesne súvisiacich fráz	KPB2	metóda
D09	Metóda na priraďovanie kľúčových fráz	KPB2	metóda
D10	Polautomatické metódy na segmentáciu právnych textov	KPB2	metóda
D16	Model dopytovania na filtrovanie právneho textu	KPB3	model
D17	Metóda vyhľadávania middleware	KPB3	metóda
D18	Metóda na hodnotenie relevantnosti	KPB4	metóda
D19	Algoritmus na prepojenie právnych textov	KPB4	algoritmus
D20	Metóda na vytváranie a popisovanie klastrov	KPB4	metóda
D21	Model na predikciu rozsudkov	KPB4	model
D22	Algoritmus na nájdenie odľahlých hodnôt	KPB4	algoritmus

Očakávaná hodnota štrnástich publikovaných metód bola dosiahnutá v plnom rozsahu. Ukazovateľ je splnený. Metódy z výstupov D05 a D22 boli publikované aj vo forme

recenzovaných vedeckých článkov — metóda na extrakciu a rozlišovanie referencií na právne predpisy a metóda detekcie anomálií v právnych textoch (kapitola 3). Zdrojové kódy všetkých metód sú spravované vo verejne prístupných repozitároch v súlade so zásadou otvoreného prístupu deklarovanou v opise projektu.

8.3 D02 - Metóda spracovania právnych textov

Údaj	Hodnota
Číslo výstupu	D02
Pracovný balík / úloha	KPB1, úloha KPB1.2
Zodpovedný partner	Wolters Kluwer SR s. r. o.
Typ	Metóda
Prístup a šírenie	verejný — publikovanie v otvorených úložiskách
Dokument s podrobným popisom	ALTAML-D02-KPB1.2-Metoda-spracovania-pravnych-textov (SK)
Dostupnosť	metodický dokument publikovaný na webovom sídle projektu

Zhrnutie: Metóda predstavuje ucelený, reprodukovateľný postup spracovania súdnych rozhodnutí pozostávajúci zo štyroch vrstiev s definovanými vstupmi, výstupmi a kontrolami: získavanie dát (crawl, API, manuálne sťahovanie), transformácia (PDF → XML → HTML, XSLT, prelinkovanie), kvalita a obohatenie (validácie, mapovanie metadát, inferencia vzťahov) a publikácia a export (CMS, dátové balíky, HTML výstupy). Pri hromadnom získavaní sú PDF dokumenty konvertované do XML skriptom využívajúcim poppler-utils a XSL transformácie, následne aplikácia doplní odkazy na právne predpisy, súdne rozhodnutia a legislatívu EÚ a dokumenty sa importujú do systému správy obsahu, kde prebehne validácia XML schémy, kontrola povinných metadát (druh, dátum, spisová značka, ECLI), konzistencie prelinkovania a anti-duplicitné pravidlá; harmonogram ťahov je riadený cronom a deduplikácia prebieha podľa hash súboru a spisovej značky. Pri manuálnom procese sa na obrazové PDF aplikuje OCR nástrojom ABBYY FineReader a zošity rozhodnutí sa segmentujú podľa titulkov, číslovania a vzorov. Dokument podrobne popisuje štyri automatizované zdroje (justice.gov.sk vrátane API endpointov, nsud.sk, nssud.sk s piatimi kategóriami rozhodovacej činnosti a ustavnysud.sk) a tri manuálne zdroje zbierok rozhodnutí. Súčasťou metódy je aj nástroj na export metadát a tvorbu dátových balíkov, ktorý na základe používateľom zadaných SQL dotazov generuje výsledkové sady s jednotnou štruktúrou metadát (Id, aspi_id, autor, spzn, druh, datum, pramen, pub_cislo, full_pramen, oblast, relation, rejstrik, rok, porc, vydani).

Väzba na ciele projektu: Výstup napĺňa špecifický cieľ 1 projektu. Metóda je určená pre akýkoľvek subjekt, ktorý má záujem o spracovanie právnych textov, a tvorí technický základ pre vznik datasetu D04.

8.4 D03 - Metódy pre kontinuálne sťahovanie

Údaj	Hodnota
Číslo výstupu	D03
Pracovný balík / úloha	KPB1, úloha KPB1.3
Zodpovedný partner	Univerzita Pavla Jozefa Šafárika v Košiciach
Typ	metóda
Prístup a šírenie	verejný — publikovanie v otvorených úložiskách
Dokument s podrobným popisom	ALTAML-D03-KPB1.3-Metody-pre-kontinualne-stahovanie (SK)
Dostupnosť	https://gitlab.science.upjs.sk/altaml/decisions-downloader

Zhrnutie: Výstupom je vysoko odolný systém pre kontinuálne sťahovanie súdnych rozhodnutí implementovaný v jazyku Python s využitím architektúry mikroslužieb a komunikácie prostredníctvom message brokera RabbitMQ. Systém automaticky získava naposledy zverejnené rozhodnutia z portálu Ministerstva spravodlivosti SR, spracúva a transformuje stiahnuté metadáta a PDF súbory do jednotného formátu JSON a ukladá ich do databázy MongoDB; vybrané atribúty sa paralelne ukladajú do relačnej databázy MySQL pre rýchle vyhľadávanie. Mikroslužbová architektúra umožňuje replikáciu jednotlivých služieb na viaceré servery a tým rozloženie záťaže, pričom každý server je identifikovaný vlastnou hodnotou premennej SERVER_NUMBER. Systém je navrhnutý ako odolný voči zlyhaniam externých aj interných komponentov: obsahuje monitorovaciu službu odosielajúcu e-mailové upozornenia na zlyhania dátového toku, administrátorské služby na export zlyhaných rozhodnutí. Celé riešenie je plne kontajnerizované pomocou Docker a Docker Compose so samostatnými compose súbormi pre databázy, RabbitMQ, dátový tok, e-mailovú monitorovaciu službu, stream a administrátorské operácie, takže nasadenie nevyžaduje manuálnu inštaláciu závislostí.

Väzba na ciele projektu: Výstup napĺňa špecifický cieľ 1 projektu a zabezpečuje priebežnú aktualizáciu dátovej základne projektu. Systém je zároveň kľúčovým komponentom architektúry opísanej vo výstupe D15.

8.5 D05 - Metóda na extrakciu referencií

Údaj	Hodnota
Číslo výstupu	D05
Pracovný balík / úloha	KPB2, úloha KPB2.1
Zodpovedný partner	Univerzita Pavla Jozefa Šafárika v Košiciach
Typ	Metóda

Údaj	Hodnota
Prístup a šírenie	verejný — publikovanie v otvorených úložiskách
Dokument s podrobným popisom	ALTAML-D05-KPB2.1-Metoda-na-extrakciu-referencii (SK)
Dostupnosť	https://gitlab.science.upjs.sk/altaml/referex

Zhrnutie: Metóda, implementovaná ako projekt ALTAML Referex, zabezpečuje automatickú extrakciu odkazov na slovenské právne predpisy, na predchádzajúce rozhodnutia slovenských súdov, na rozhodnutia Európskeho súdu pre ľudské práva a Súdneho dvora Európskej únie a na rozhodnutia Úradu na ochranu osobných údajov SR. Kombinuje pravidlami riadené prístupy cez regulárne výrazy, Levenshteinovu vzdialenosť a fuzzy vyhľadávanie s rozpoznávaním pomenovaných entít (NER) založeným na modeli SlovakBERT. Riešenie je nasaditeľné ako REST API služba postavená s dvoma endpointmi: /v1/referex/laws pre extrakciu odkazov na právne predpisy a /v1/referex/dockets pre extrakciu odkazov na súdne rozhodnutia. Podpornú dátovú vrstvu tvorí databáza MongoDB so šiestimi kolekciami (law_validities, law_aliases, law_aliases_lemmatized, codebook_courts, zakony_aliases, dict_lemmatized_laws), ktoré umožňujú identifikáciu zákona aj vtedy, keď je v texte použitá iba skratka, priradenie správnej časovej verzie predpisu a spoľahlivé rozpoznanie súdu uvedeného v rozhodnutí. Na lematizáciu vstupného textu služba využíva Tvaroslovník API. Dôsledkom tejto metódy je, že umožňuje vytvárať hypertextové a sémantické prepojenia medzi právnymi dokumentmi.

Väzba na ciele projektu: Výstup napĺňa špecifický cieľ 2 projektu a podporuje dlhodobý cieľ budovania prepojeného a vyhľadateľného ekosystému slovenských a európskych právnych textov. Metóda rozlišovania referencií bola publikovaná aj vo forme recenzovaného vedeckého článku na konferencii JURIX 2025 (kapitola 3).

8.6 D06 - Metóda na extrakciu rozsudkov

Údaj	Hodnota
Číslo výstupu	D06
Pracovný balík / úloha	KPB2, úloha KPB2.2
Zodpovedný partner	Univerzita Pavla Jozefa Šafárika v Košiciach
Typ	metóda
Prístup a šírenie	verejný — publikovanie v otvorených úložiskách
Dokument s podrobným popisom	ALTAML-D06-KPB2.2-Metoda-na-extrakciu-rozsudkov (SK)
Dostupnosť	https://gitlab.science.upjs.sk/altaml/verdict-extraction

Zhrnutie: Výstupom je hybridný extraktor údajov z trestnoprávnych rozhodnutí, ktorý kombinuje pravidlové spracovanie textu, heuristické metódy spracovania prirodzeného jazyka a veľký jazykový model LLaMA 3.3 lokálne nasadený prostredníctvom platformy Ollama. Systém extrahuje skutkovú vetu, informáciu o oslobodení spod obžaloby, uložené tresty, poľahčujúce okolnosti podľa § 36 Trestného zákona, priťažujúce okolnosti podľa § 37 Trestného zákona a dôkazné prostriedky, ktoré kategorizuje na výsluchy, listinné dôkazy, znalecké posudky, vecné dôkazy a záznamy a ostatné dôkazy. Skutková veta sa extrahuje v čo najvernejšej podobe priamo z textu rozhodnutia, bez sumarizácie alebo preformulovania. Pred samotnou extrakciou systém jazykovým modelom overí počet obžalovaných a rozhodnutia s viacerými páchateľmi automaticky vyradí, čím sa znižuje riziko nesprávnej atribúcie údajov jednotlivým osobám. Identifikácia trestných činov prebieha vo viacerých stupňoch: heuristický výber kandidátov na základe lexikálneho prekryvu skutkovej vety s popismi trestných činov (s tokenizáciou, odstránením stop-slov a špecializovanými heuristikami pre časté skutkové situácie) a následný reranking jazykovým modelom, ktorý je striktným promptom obmedzený výlučne na uzavretú databázu slovenských trestných činov. Pri zlyhaní modelu sa aktivuje pravidlový fallback mechanizmus. Na spracovanie rozsiahlych korpusov sa používa streamingový parser ijson, výsledky sa ukladajú priebežne. Odporúčaná konfigurácia je minimálne 32 GB RAM s podporou GPU (CUDA).

Väzba na ciele projektu: Výstup napĺňa špecifický cieľ 2 projektu a zároveň priamo reaguje na požiadavku koncových používateľov na stručné zhrnutia súdnych rozhodnutí identifikovanú vo výstupe D01. Extrahované atribúty tvoria vstup pre výstup D21.

8.7 D07 - Metóda na extrakciu procesne súvisiacich fráz

Údaj	Hodnota
Číslo výstupu	D07
Pracovný balík / úloha	KPB2, úloha KPB2.3
Zodpovedný partner	Univerzita Pavla Jozefa Šafárika v Košiciach
Typ	Metóda
Prístup a šírenie	verejný — publikovanie v otvorených úložiskách
Dokument s podrobným popisom	ALTAML-D07-KPB2.3-Metoda-na-extrakciu-procesne-suvisiacich-fraz (SK)
Dostupnosť	https://gitlab.science.upjs.sk/altaml/process-related-phrases

Zhrnutie: Metóda predstavuje hybridný extraktor procesných právnych udalostí zameraný na automatickú identifikáciu procesných úkonov a výsledkov konaní zo slovenských súdnych rozhodnutí, konkrétne na zamietnutie odvolania, úspešné odvolanie, zastavenie konania, zrušenie rozhodnutia a vrátenie veci na ďalšie konanie. Extrakcia je obmedzená na vopred definovanú ontológiu procesných kategórií, v ktorej každá kategória obsahuje množinu

reprezentatívnych formulácií slúžiacich ako referenčné prototypy. Spracovanie prebieha viacvrstvovo: po normalizácii textu a rozdelení na vety sa vykoná heuristická filtrácia významových viet podľa prítomnosti procesných kľúčových slov, dĺžky vety a výskytu právnej terminológie typickej pre výrokové časti; v prvej fáze extrakcie sa aplikujú regulárne výrazy na presnú identifikáciu štandardizovaných formulácií, a ak pravidlový prístup neprinesie výsledok, systém využije lokálne nasadený jazykový model LLaMA 3.3 cez Ollama riadený striktným slovenským promptom. Extrahované frázy sa následne prevedú do vektorovej reprezentácie modelom SlovakBERT doladeným na úlohu sémantickej podobnosti viet a klasifikujú sa podľa najvyššej kosínovej podobnosti s referenčnými prototypmi pri uplatnení minimálneho prahu. Po klasifikácii nasleduje deduplikácia podľa kombinácie kategórie a normalizovanej podoby frázy a zoradenie podľa skóre. Výstupom je JSON štruktúra, ktorá pre každú udalosť uvádza kategóriu, extrahovanú frázu, skóre podobnosti, referenčný prototyp, pôvodnú vetu rozhodnutia a metódu extrakcie.

Väzba na ciele projektu: Výstup napĺňa špecifický cieľ 2 projektu. Kombinácia pravidlového a sémantického prístupu umožňuje spracovať aj jazykovo variabilné formulácie typické pre slovenské súdne rozhodnutia a vytvárať štruktúrované databázy procesných udalostí.

8.8 D09 - Metóda na priraďovanie kľúčových fráz

Údaj	Hodnota
Číslo výstupu	D09
Pracovný balík / úloha	KPB2, úloha KPB2.5
Zodpovedný partner	Univerzita Pavla Jozefa Šafárika v Košiciach
Typ	Metóda
Prístup a šírenie	verejný — publikovanie v otvorených úložiskách
Dokument s podrobným popisom	ALTAML-D09-KPB2.5-Metoda-pridelovania-klucovych-pojmov (SK)
Dostupnosť	https://gitlab.science.upjs.sk/altaml/keyphrase-extractor

Zhrnutie: Metóda predstavuje hybridný extraktor právnych kľúčových pojmov zo slovenských súdnych rozhodnutí, ktorý kombinuje štatistické spracovanie textu, sémantické vektorové reprezentácie a veľký jazykový model. Extrakcia je obmedzená na uzavretú množinu právnych pojmov pochádzajúcu z výstupu D08, čím sa zabezpečuje terminologická presnosť a minimalizuje riziko generovania neustálených výrazov. Skórovanie kandidátnych pojmov má štyri zložky: TF-IDF váženie vypočítané výhradne nad množinou kandidátnych pojmov (ostatný lexikálny obsah dokumentu sa ignoruje a slovenské stop-slová sú explicitne odstránené), kosínová podobnosť medzi embeddingom pojmu a embeddingom celého dokumentu vytvorenými modelom SlovakBERT doladeným na sémantickú podobnosť viet, modifikácia skóre podľa výskytu pojmu v referenčných textoch (odkazovaných rozhodnutiach a citovaných právnych predpisoch) normalizovaná počtom referenčných textov, a napokon lineárna

kombinácia štatistickej a sémantickej zložky s regulovateľným váhovým parametrom. Embedding kandidátneho pojmu kombinuje samotný názov pojmu s jeho definíciami z databázy D08. Voliteľne je možné uplatniť princíp maximálnej marginálnej relevancie (MMR), ktorý zabezpečí významovú rozmanitosť vybraných pojmov. Finálnu selekciu vykonáva lokálne nasadený model LLaMA 3.3 cez Ollama, ktorý odstraňuje synonymické a duplicitné výrazy a uprednostňuje právne ustálené pojmy, čím simuluje expertné rozhodovanie právnika.

Väzba na ciele projektu: Výstup napĺňa špecifický cieľ 2 projektu a reaguje na požiadavku koncových používateľov na extrakciu kľúčových slov nad rozhodnutiami identifikovanú vo výstupe D01. Extrahované kľúčové pojmy sú zároveň vstupom pre výstupy D16, D18 a D19. Prístup kombinujúci kľúčové frázy a referencie pri vyhľadávaní dokumentov bol publikovaný vo vedeckom článku na workshope JURISIN 2025, kde dosiahol strednú priemernú presnosť (mAP) 92,85 %.

8.9 D10 - Polautomatické metódy na segmentáciu právnych textov

Údaj	Hodnota
Číslo výstupu	D10
Pracovný balík / úloha	KPB2, úloha KPB2.6
Zodpovedný partner	Univerzita Pavla Jozefa Šafárika v Košiciach
Typ	Metóda
Prístup a šírenie	verejný — publikovanie v otvorených úložiskách
Dokument s podrobným popisom	ALTAML-D10-KPB2.6-Poloautomatizovana-metoda-na-segmentaciu-pravnych-textov (SK)
Dostupnosť	https://gitlab.science.upjs.sk/altaml/legal-text-segmentation

Zhrnutie: Výstupom je hybridný segmentačný a klasifikačný model, ktorý automaticky rozdeľuje slovenské súdne rozhodnutia na logické obsahové časti a každej časti priraďuje právne relevantný typ segmentu. Systém pracuje viacúrovňovo. Najprv vykoná hrubú segmentáciu na úrovni tematických blokov analýzou sémantických vzdialeností medzi okolitými vetami, ktorých vektorové reprezentácie vytvára model SlovakBERT doladený na sémantickú podobnosť viet; kandidátne hranice sa určujú podľa lokálnych maxím skóre tematickej zmeny a dopĺňajú sa makrohranicami identifikovanými pomocou štruktúrnych markerov typických pre slovenské rozhodnutia. Segmenty sa následne klasifikujú do desiatich právnych kategórií — Predmet_sporu, Vyrokova_cast, Posudenie_sudu, Skutkove_okolnosti, Pravny_ramec, Dokazy, Predchadzajuce_konanie_vo_veci, Zalobca_Prokurator, Zalovany_Obzalovany a Dalsie. Klasifikácia kombinuje znakové TF-IDF n-gramy typu char_wb, ktoré robustne zachytávajú morfologické varianty slovenského právneho jazyka, sémantické skórovanie voči prototypovým opisom tried a rozsiahlu množinu pravidlových heuristík zohľadňujúcich právne formulácie, procesné výrazy, odkazy na predpisy, identifikáciu strán

konania, výrokové formulácie, súdne odôvodnenie a štruktúrnu pozíciu segmentu. Nasleduje sekvenčný postprocessing odstraňujúci nekonzistencie medzi susediacimi segmentmi a druhá úroveň klasifikácie — vetný relabeling využívajúci embeddingovú podobnosť, lokálny kontext susedných viet a sekvenčné pravidlá. Výstupom je štruktúrovaná JSON reprezentácia dokumentu a analytické CSV súbory so štatistikami relabelingu a prechodovými maticami medzi triedami.

Väzba na ciele projektu: Výstup napĺňa špecifický cieľ 2 projektu. Segmentácia umožňuje zúžiť oblasť extrakcie relevantných údajov a filtrovať rozhodnutia podľa umiestnenia odkazu alebo pojmu v konkrétnej časti rozhodnutia, čo priamo využívajú výstupy D16, D17, D18 a D19.

8.10 D16 - Model dotazovania na filtrovanie právneho textu

Údaj	Hodnota
Číslo výstupu	D16
Pracovný balík / úloha	KPB3, úloha KPB3.3
Zodpovedný partner	Univerzita Pavla Jozefa Šafárika v Košiciach
Typ	Model
Prístup a šírenie	verejný — publikovanie v otvorených úložiskách
Dokument s podrobným popisom	ALTAML-D16-KPB3.3-Dopytovaci-model-pre-vyhladavanie-pravnych-textov (SK)
Dostupnosť	dokumentácia dopytovacieho modelu; indexy nad Elasticsearch podľa architektúry D15

Zhrnutie: Výstup definuje dopytovací model postavený na Elasticsearch Query DSL, ktorý umožňuje kombinovať fulltextové vyhľadávanie v textoch rozhodnutí a právnych predpisov, vyhľadávanie podľa extrahovaných odkazov na predpisy a na iné rozhodnutia, vyhľadávanie podľa automaticky identifikovaných kľúčových pojmov, filtrovanie podľa kategorizovaných atribútov a sémantické vyhľadávanie pomocou vektorových reprezentácií, vrátane hybridných dopytov. Model používa samostatné indexy pre súdne rozhodnutia a pre právne predpisy. Pre fulltextové polia je definovaný vlastný analyzér slovak_lemmatizer založený na slovenskom Hunspell slovníku, ktorý umožňuje dosiahnuť zhodu aj medzi rôznymi gramatickými tvarmi toho istého slova. Index súdnych rozhodnutí obsahuje sedemnásť atribútov s priradenými dátovými typmi — text pre fulltext, keyword pre presné filtrovanie a agregácie (ECLI, spisová značka, forma rozhodnutia, oblasť a podoblasť právnej úpravy, kľúčové pojmy, názov a typ súdu), date pre rozsahové dopyty a nested pre vnorené štruktúry odkazov. Keďže rozhodnutia môžu byť rozsiahle, vektorová reprezentácia sa nevytvára pre celý dokument, ale pre menšie textové časti (chunky), z ktorých každý nesie informáciu o časti rozhodnutia, z ktorej pochádza (hlavička, úvod, výrok, odôvodnenie, poučenie); to umožňuje obmedziť sémantické vyhľadávanie napríklad iba na odôvodnenie. Embeddingy sú 384-rozmerné, generované

modelom paraphrase-multilingual-MiniLM-L12-v2, a porovnávajú sa kosínovou podobnosťou prostredníctvom knn dopytov s približným vyhľadávaním najbližších susedov. Dokument obsahuje osem komentovaných ukážok dopytov vrátane hybridného vyhľadávania, ktoré váženou kombináciou (0,7 ku 0,3) spája fulltextové skóre BM25 s kosínovou podobnosťou embeddingov. Pri právnych predpisoch sa text nedelí na chunky a používa sa jedna vektorová reprezentácia celého predpisu spolu s filtrovaním podľa obdobia platnosti.

Väzba na ciele projektu: Výstup napĺňa špecifický cieľ 3 projektu a priamo reaguje na najsilnejšiu požiadavku koncových používateľov — rýchle a presné vyhľadávanie — identifikovanú vo výstupe D01. Nadväzuje na architektúru D15 a tvorí základ pre používateľskú vrstvu D17.

8.11 D17 - Metóda vyhľadávania middleware

Údaj	Hodnota
Číslo výstupu	D17
Pracovný balík / úloha	KPB3, úloha KPB3.4
Zodpovedný partner	Univerzita Pavla Jozefa Šafárika v Košiciach
Typ	Metóda
Prístup a šírenie	verejný — publikovanie v otvorených úložiskách
Dokument s podrobným popisom	ALTAML-D17-KPB3.4-Metoda-vyhladavania-middleware (SK)
Dostupnosť	https://gitlab.science.upjs.sk/altaml/middleware-search-method

Zhrnutie: Metóda rieši situáciu, keď používateľ pozná iba všeobecnú tému, ktorú chce vyhľadávať, ale nevie hneď formulovať všetky súvisiace pojmy a právne ustanovenia vedúce k relevantným výsledkom. Vstupom je preto aj všeobecne formulovaná téma, napríklad „kyberkriminalita“. Metóda ju najprv prevedie modelom paraphrase-multilingual-MiniLM-L12-v2 na 384-rozmerný embedding a pomocou knn dopytov vyhľadá sémanticky súvisiace chunky súdnych rozhodnutí a právne ustanovenia v dvoch samostatných Elasticsearch indexoch. Rozhodnutia sú v indexe uložené po chunkoch s maximálnou veľkosťou 400 znakov, generovaných nad hlavnými časťami dokumentu (úvod, výrok, odôvodnenie, poučenie). Právne ustanovenia sú uložené hierarchicky na úrovni paragrafu, odseku aj písmena, pričom konkrétnejšie dokumenty vždy nesú nadradený kontext, aby zostali významovo úplné; pri vyhľadávaní sa následne vykoná hierarchická deduplikácia, ktorá pri súčasnom výskyte paragrafu aj jeho odseku ponechá konkrétnejšie ustanovenie. Z vyhľadaných výsledkov sa vytvoria dva textové kontexty, ktoré vstupujú do dvoch samostatných volaní jazykového modelu riadených systémovým a používateľským promptom a obmedzených JSON schémou. Prvé volanie navrhne 5 až 10 slovenských vyhľadávacích pojmov odvodených výlučne z

vyhľadaných rozhodnutí (napríklad „IP adresa“, „počítačová sieť“, „internetová stránka“), druhé vyberie identifikátory relevantných právnych ustanovení (napríklad § 247a alebo § 360b Trestného zákona) výlučne zo zoznamu predložených kandidátov. Výsledkom je jeden JSON objekt s poľami search_phrases a legal_provisions, z ktorých si používateľ vyberá hodnoty pre ďalšie vyhľadávanie. Metóda bola implementovaná ako spustiteľný prototyp vrátane demo skriptu demo_cybercrime.py.

Väzba na ciele projektu: Výstup napĺňa špecifický cieľ 3 projektu. Znižuje bariéru vstupu pri vyhľadávaní právnych informácií pre používateľov, ktorí nedisponujú detailnou znalosťou terminológie a právnych ustanovení v danej oblasti, čo bolo jedným z opakujúcich sa zistení prieskumu D01.

8.12 D18 - Metóda na hodnotenie relevantnosti

Údaj	Hodnota
Číslo výstupu	D18
Pracovný balík / úloha	KPB4, úloha KPB4.1
Zodpovedný partner	Univerzita Pavla Jozefa Šafárika v Košiciach
Typ	Metóda
Prístup a šírenie	verejný — publikovanie v otvorených úložiskách
Dokument s podrobným popisom	ALTAML-D18-KPB4.1-Metoda-na-hodnotenie-relevantnosti (SK)
Dostupnosť	https://gitlab.science.upjs.sk/altaml/document-similarity

Zhrnutie: Metóda určuje mieru podobnosti dvoch slovenských súdnych rozhodnutí kombináciou piatich komponentov: globálnej sémantickej podobnosti dokumentov, podobnosti jednotlivých textových segmentov, zhody odkazov na súvisiace rozhodnutia, zhody citovaných právnych predpisov a podobnosti extrahovaných kľúčových pojmov. Každé rozhodnutie sa najprv rozdelí na chunky na úrovni odsekov agregovaných do približne rovnako veľkých celkov, čo umožňuje analyzovať podobnosť nielen na úrovni celého dokumentu, ale aj medzi konkrétnymi časťami právnej argumentácie. Podstatným krokom je filtrovanie procesného a „šablónového“ textu — označení súdov, procesných poučení, štandardizovaných formulácií a opakujúcich sa právnických fráz — ktoré znižuje vplyv všeobecného procesného jazyka na výsledné skóre a zvyšuje dôraz na vecný obsah rozhodnutí. Vektorové reprezentácie dokumentov aj chunkov vytvára model SlovakBERT doladený na sémantickú podobnosť viet. Podobnosť chunkov sa vyhodnocuje symetricky v oboch smeroch a do výsledku vstupujú len najrelevantnejšie zhody nad definovaným prahom. Zhoda referencií a citovaných predpisov sa počíta Jaccardovou podobnosťou množín identifikátorov a podobnosť kľúčových pojmov váženým prekryvom zohľadňujúcim ich relatívnu významnosť. Finálne skóre je lineárnou kombináciou uvedených komponentov, pričom najväčšiu váhu má

podobnosť chunkov, ktorá najlepšie reprezentuje vecnú a argumentačnú podobnosť. Výstupom nie je iba numerické skóre, ale aj vysvetliteľné informácie — čiastkové skóre jednotlivých metrík, zoznam spoločných právnych referencií a predpisov, spoločné kľúčové pojmy a najpodobnejšie textové segmenty.

Väzba na ciele projektu: Výstup napĺňa špecifický cieľ 4 projektu (identifikácia vzťahov medzi právnymi textami). Vysvetliteľnosť výsledkov je pritom podmienkou akceptovateľnosti metódy v právnej praxi, čo potvrdila aj spätná väzba právnych výskumných tímov zaznamenaná vo výstupe D23.

8.13 D19 - Algoritmus na prepojenie právnych textov

Údaj	Hodnota
Číslo výstupu	D19
Pracovný balík / úloha	KPB4, úloha KPB4.2
Zodpovedný partner	Univerzita Pavla Jozefa Šafárika v Košiciach
Typ	algoritmus
Prístup a šírenie	verejný — publikovanie v otvorených úložiskách
Dokument s podrobným popisom	ALTAML-D19-KPB4.2-Algoritmus-pre-prepajanie-pravnych-textov (SK)
Dostupnosť	https://gitlab.science.upjs.sk/altaml/document-relevance-extractor

Zhrnutie: Algoritmus nadväzuje na metriku podobnosti vyvinutú vo výstupe D18 a aplikuje ju na celý korpus: po zadaní zdrojového dokumentu porovná referenčné rozhodnutie so všetkými ostatnými rozhodnutiami v dátovej sade, vypočíta skóre podobnosti pre každú dvojicu a vráti zoradený zoznam identifikátorov najpodobnejších rozhodnutí. Rovnako ako pri D18 sa uplatňuje hybridný prístup kombinujúci sémantickú podobnosť textu, podobnosť obsahových častí dokumentu, zhodu právnych referencií, zhodu citovaných právnych predpisov a podobnosť extrahovaných kľúčových pojmov, pričom pred výpočtom prebieha segmentácia dokumentov na chunky a filtrovanie procesného a formulárového textu. Vektorové reprezentácie vytvára model SlovakBERT doladený na sémantickú podobnosť viet, porovnanie prebieha kosínovou podobnosťou a zhoda referencií a predpisov sa vyhodnocuje Jaccardovou podobnosťou. Výsledkom je sieť prepojených právnych dokumentov reflektujúca vecnú a významovú súvislosť medzi rozhodnutiami, ktorá umožňuje identifikovať napríklad „rodinu rozhodnutí“ v tej istej veci alebo rozhodnutia s podobnou právnou argumentáciou či skutkovým základom. Systém je implementovaný v jazyku Python s využitím knižníc Sentence Transformers, scikit-learn a NumPy a podporuje GPU akceleráciu (CUDA) pre efektívne spracovanie rozsiahlych korpusov.

Väzba na ciele projektu: Výstup napĺňa špecifický cieľ 4 projektu. Zoradený zoznam relevantných dokumentov predstavuje odporúčací nástroj, ktorý pomáha právnikovi identifikovať rozhodnutia hodné hlbšieho manuálneho skúmania.

8.14 D20 - Metóda na vytváranie a popisovanie klastrov

Údaj	Hodnota
Číslo výstupu	D20
Pracovný balík / úloha	KPB4, úloha KPB4.3
Zodpovedný partner	Univerzita Pavla Jozefa Šafárika v Košiciach
Typ	metóda
Prístup a šírenie	verejný — publikovanie v otvorených úložiskách
Dokument s podrobným popisom	ALTAML-D20-KPB4.3-Metoda-na-vytvaranie-a-popisovanie-klastrov (SK)
Dostupnosť	https://github.com/simonHorvat/GCDC

Zhrnutie: Metóda predstavuje grafovo orientovaný kategorizátor súdnych rozhodnutí implementovaný v jazyku Python s využitím knižnice NetworkX. Rozhodnutia, právne odkazy a kľúčové pojmy tvoria uzly multipartitného grafu, ktorého hrany zachytávajú viacrozmerný charakter právnych textov. Príprava dát zahŕňa extrakciu právnych odkazov pomocou regulárnych výrazov a metrík podobnosti vrátane Levenshteinovej vzdialenosti, prípravu množiny kandidátnych kľúčových pojmov a vytvorenie vektorových reprezentácií textov, ich referencií aj kandidátnych pojmov (TF-IDF, resp. model bge-m3). Skóre kandidátneho pojmu pre daný dokument vzniká kombináciou TF-IDF váženia a kosínovej podobnosti, pričom sa zlučuje skóre zo samotného textu so skóre z jeho referencií a následne sa aplikuje prahovanie. Váha hrany medzi rozhodnutím a právnym odkazom sa počíta ako podiel lokálneho počtu citácií daného odkazu v konkrétnom rozhodnutí a globálneho počtu citácií toho istého odkazu v celej dátovej sade, pričom všetky váhy sú normalizované na úrovni rozhodnutia tak, aby ich súčet bol rovný jednej. Významné, ale zriedkavé citácie tak získavajú vyššiu váhu než často sa vyskytujúce. Latentné právne oblasti sa identifikujú detekciou komunít v dvoch fázach: presúvaním uzlov medzi komunitami s cieľom maximalizovať prírastok modularity a následným zhlukovaním komunít do superuzlov s rekurzívnym opakovaním procesu. Metóda bola experimentálne overená na tisíc súdnych rozhodnutiach z oblasti trestného práva, kde viedla k identifikácii štrnástich samostatných komunít s hodnotou modularity 0,53; dokumentácia uvádza aj konkrétne príklady klastrov s ich kľúčovými právnymi odkazmi a kľúčovými slovami.

Väzba na ciele projektu: Výstup napĺňa špecifický cieľ 4 projektu (identifikácia skupín právnych textov). Popis klastrov prostredníctvom reprezentatívnych právnych odkazov a kľúčových slov umožňuje rýchlejšie filtrovanie a podporuje analýzu vzorcov v rozhodovacej praxi.

8.15 D21 - Model na predikciu rozsudkov

Údaj	Hodnota
Číslo výstupu	D21
Pracovný balík / úloha	KPB4, úloha KPB4.4
Zodpovedný partner	Univerzita Pavla Jozefa Šafárika v Košiciach
Typ	model
Prístup a šírenie	verejný — publikovanie v otvorených úložiskách
Dokument s podrobným popisom	ALTAML-D21-KPB4.4-Model-pre-predikciu-verdiktov (SK)
Dostupnosť	https://gitlab.science.upjs.sk/altaml/verdict-prediction

Zhrnutie: Model predstavuje hybridný analytický pipeline pre automatizovanú klasifikáciu výsledku trestného konania a následnú extrakciu uložených trestov. Architektúra je trojvrstvová. Vrstva Tier 0 identifikuje na základe explicitných paragrafových odkazov prípady, ktoré boli zastavené alebo prerušené, a vyradí ich z ďalšieho spracovania. Vrstva Tier 1 aplikuje deterministické právne pravidlá založené na výskyte paragrafov signalizujúcich oslobodenie, odsúdenie a dominantných odsudzujúcich ustanovení; ak pravidlá vedú k jednoznačnému výsledku, rozhodnutie je klasifikované bez ďalšieho spracovania. Vrstva Tier 2 sa aktivuje len pri neurčitom výsledku a využíva prístup Retrieval-Augmented Generation v ktorom sa pre analyzovaný prípad vypočíta embedding modelom BGE-M3 zahŕňajúci skutkový opis, identifikované trestné činy, poľahčujúce a priťažujúce okolnosti a dôkazy, kosínovou podobnosťou sa vyhľadajú najpodobnejšie historické prípady a tie sa vložia ako few-shot príklady do promptu lokálne nasadeného modelu LLaMA 3.3, ktorý vykoná finálne rozhodnutie. Pri odsudzujúcich rozhodnutiach systém následne extrahuje typ trestu, jeho mieru, podmienku a kategóriu výkonu a doplňujúce detaily. Konzistentnosť výstupov je zabezpečená striktne definovanými slovenskými promptami, obmedzením výstupu na JSON štruktúry, uzavretou množinou definovaných typov trestov a inferenciou s nulovou teplotou. Embeddingy sa kešujú v lokálnej JSON databáze a výsledky sa ukladajú priebežne, čím je pipeline odolný voči prerušeniu výpočtu.

Väzba na ciele projektu: Výstup napĺňa špecifický cieľ 4 projektu a zároveň priamo napĺňa požiadavku koncových používateľov na nástroj na predikciu výsledku právneho sporu, ktorá bola jedným z troch kľúčových zistení prieskumu D01. Model nadväzuje na atribúty extrahované vo výstupe D06.

8.16 D22 - Algoritmus na nájdenie odľahlých hodnôt

Údaj	Hodnota
Číslo výstupu	D22
Pracovný balík / úloha	KPB4, úloha KPB4.5
Zodpovedný partner	Univerzita Pavla Jozefa Šafárika v Košiciach
Typ	algoritmus
Prístup a šírenie	verejný — publikovanie v otvorených úložiskách
Dokument s podrobným popisom	ALTAML-D22-KPB4.5-Algoritmus-na-najdenie-odlahlych-hodnot (SK)
Dostupnosť	https://gitlab.science.upjs.sk/altaml/method-for-outlier-detection

Zhrnutie: Algoritmus identifikuje netypické právne rozhodnutia metódou učenia bez učiteľa Isolation Forest a získané výsledky interpretuje pomocou metódy SHAP. Postup má tri kroky. Príprava dát transformuje právne atribúty do numerickej reprezentácie — dopĺňa chýbajúce hodnoty a vytvára ich indikátory, normalizuje textové atribúty, prevádza binárne kategórie do numerickej podoby, dopĺňa hodnotu „unknown“ pre chýbajúce kategórie, aplikuje one-hot encoding na kategorické atribúty a odstraňuje atribúty s konštantnou hodnotou. Model Isolation Forest, ktorý pracuje na princípe náhodného rozdeľovania dátového priestoru, následne pre každé rozhodnutie vypočíta skóre anomálnosti, indikátor netypickosti podľa interného prahu daného parametrom kontaminácie nastaveným na 10 % a poradie rozhodnutí podľa miery odchýlky. Interpretačná vrstva postavená na SHAP umožňuje vysvetliť, prečo bolo konkrétne rozhodnutie označené ako anomálne: waterfall grafy zobrazujú príspevok jednotlivých atribútov k výslednému skóre na úrovni jedného rozhodnutia, graf priemernej absolútnej SHAP hodnoty a beeswarm vizualizácia poskytujú globálny pohľad na dôležitosť a smer vplyvu atribútov naprieč celou dátovou sadou a heatmapa SHAP interakcií zachytáva silu vzájomného pôsobenia dvojíc atribútov. Metóda bola experimentálne overená na piatich dátových sadách súdnych rozhodnutí manuálne anotovaných právnikmi z oblastí ako zmluvná pokuta či nekalá súťaž, čím bola overená jej robustnosť naprieč rôznymi právnymi doménami. Pre každú dátovú sadu sa generuje samostatný výstupný priečinok s desiatimi typmi výstupov (CSV so skóre a poradím, SHAP hodnoty, globálna dôležitosť atribútov, interakčné dvojice, sumarizačné súbory a vizualizácie) a naprieč všetkými sadami aj spoločný sumarizačný súbor.

Väzba na ciele projektu: Výstup napĺňa špecifický cieľ 4 projektu (identifikácia právnych textov, ktoré sa odlišujú od ostatných). Metóda detekcie anomálií v právnych textoch bola publikovaná aj vo forme recenzovaného vedeckého článku na konferencii JURIX 2025 (kapitola 3) a podľa spätnej väzby právnych výskumných tímov zaznamenanej vo výstupe D23 patrí spolu s deskriptívnou štatistickou analýzou k najužitočnejším analytickým nástrojom projektu.