Extrakcia právnych odkazov pomocou modelov transformérov

Vedúci práce: RNDr. Peter Gurský, PhD.

Konzultant: RNDr. Dávid Varga

Študent: Bc. Nicol Fedurcová

UNIVERZITA PAVLA JOZEFA ŠAFÁRIKA V KOŠICIACH

Prečo je extrakcia právnych odkazov dôležitá?

- Využiteľnosť v súdnej vede :
  - Nájdenie rozhodnutí, týkajúce sa podobných prípadov
  - Skúmanie vzťahov medzi rozhodnutiami
- Aplikácia v systéme na vyhľadávanie súdnych rozhodnutí
- Zjednodušenie automatizovaného spracovania rozhodnutí

Výzvy pri vyhľadávaní a anotovaní právnych odkazov

HLAVNÝ PROBLÉM: súdne rozhodnutia sú písané prirodzeným jazykom

- Odkazy nemajú jednotnú štruktúru
- Používanie skratiek, značiek, skratkových slov, ktoré nemusia byť vždy vysvetlené
- Každý sudca má vlastný štýl písania rozhodnutia
- Štruktúra rozhodnutí je naprieč rôznymi krajinami rôzna

Ciele práce

1. Vypracovať prehľad metód extrakcie z textov s využitím transformerov a prehľad transformer-modelov vhodných pre slovenské právne texty.
2. Identifikovať a analyzovať dôvody neúspešných extrakcií v existujúcom pravidlovom systéme extrahujúcom odkazy na súdne rozhodnutia a právne predpisy.
3. Navrhnúť a implementovať rozšírenie existujúceho riešenia s využitím transformer modelu.
4. Porovnať úspešnosť novej metódy s existujúcim pravidlovým systémom na rozšírenom anotovanom datasete právnych textov.

vykonaná aktivita

Čo sme urobili

- Prehľad použiteľných jazykových modelov
- Hľadanie príčin zlyhania extrakcie odkazov pôvodnou metódou
- Vytvorenie metódy na nájdenie rozhodnutí s viacerými možnosťami výberu pri aliase
- Anotovanie rozhodnutí s viacerými možnosťami výberu pri aliase, Vytvorenie prvej verzie Choice checkera
- Modifikovanie metódy na nájdenie rozhodnutí s viacerými možnosťami výberu pri aliase

Čo sme urobili

- Vytvorenie metódy na nájdenie a preformátovanie znenia konkrétneho odkazu
- Vytvorenie test makera, ktorého výstup bude vstupom pre nový choice checker
- Vytvorenie novej (zjednodušenej) verzie choice checkera (kontrolóra viacerých možností pri aliase – bez kontroly existencie odkazu)
- Zostavenie a manuálne anotovanie testovacej sady na porovnanie choice checkerov s rôznymi parametrami
- Porovanie výsledkov rôznych parametrov choice checkera

Hľadanie príčin zlyhania extrakcie odkazov pôvodnou metódou

Príklad A)

Súd podľa § 101 ods. 2 O.s.p. konal a rozhodol v neprítomnosti navrhovateľa a jeho právneho zástupcu, ktorí svoju neúčasť na pojednávaní ospravedlnili, súhlasili s pojednávaním v ich neprítomnosti.

```
_id: "O.s.p"
law_ids: Array
  0: "99/1963"
```

Príklad B)

§ 62 ods. 1 zákona č. 36/2005 Z. z. o rodine a o zmene a doplnení niektorých zákonov - ďalejlen \"ZoR\"

Podľa § 78 ods. 2 ZoR ak dôjde k zrušeniu alebo zníženiu výživného pre maloleté dieťa

```
_id: "ZoR"
law_ids: Array
  0: "94/1963"
  1: "530/2003"
  2: "36/2005"
```

Príklad C)

Uvedené ust. § 23 ZoR predpokladá objektívnuexistenciu rozvratu vzťahov bez ohľadu na subjektívne

Testovací dataset

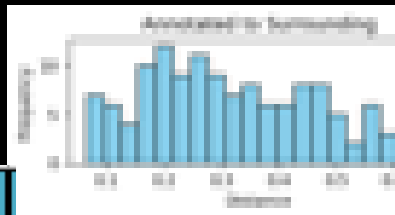
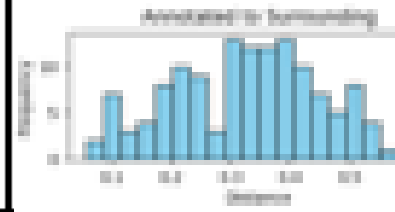
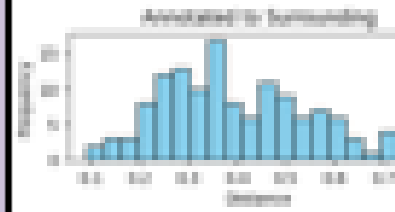
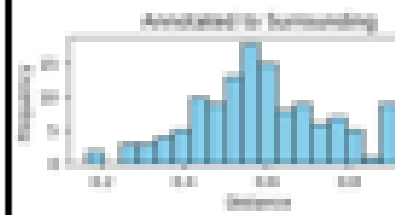
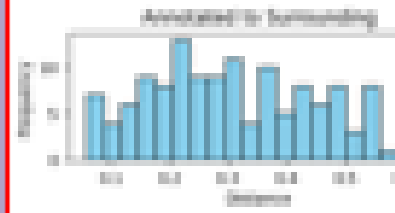

```
{
    "url": "/SK/ZZ/1977/61/19930822.html#paragraf-127.odsek-4",
    "law": "61/1977",
    "version": "19930822",
    "fragment": "paragraf-127.odsek-4",
    "section": 4,
    "letter": null,
    "start": 5937,
    "end": 5953,
    "paragraph": "127",
    "zakonnik_start": 5949,
    "zakonnik_end": 5953,
    "extracted_from": "global from db",
    "alias_text": "",
    "aliases": "ZoCP",
    "alias_options": [
        "61/1977",
        "526/1990",
        "566/2001",
        "326/2005",
        "8/2009"
    ],
    "annotated_url": "/SK/ZZ/2001/566/20130610.html#paragraf-127.odsek-4",
    "existing_options": [
        {
            "zakonnik": "566/2001",
            "provision_text": "566/2001, o cenných papieroch a investičných službách a o zmene a doplnení niektorých zákonov      (zákon o cenných papieroch), och
            "version": "20130610.html"
        },
        {
            "zakonnik": "8/2009",
            "provision_text": "8/2009, o cestnej premávke a o zmene a doplnení niektorých zákonov, evidencia vozidiel, evidovanie vozidiel a evidenčné čísla, evid
            "version": "20160101.html"
        }
    ],
    "surrounding_text": "neskorších predpisov /ZoCP/  za porušenie ustanovenia § 129 ods. 3 v spojení             sustanovením § 128 ods.1 písmeno b/ c/ h/ ZoCP
}
```

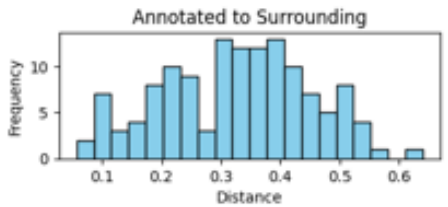
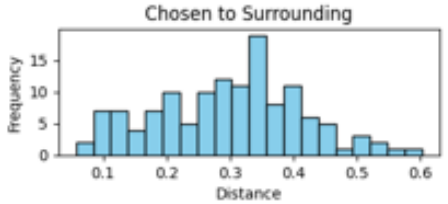
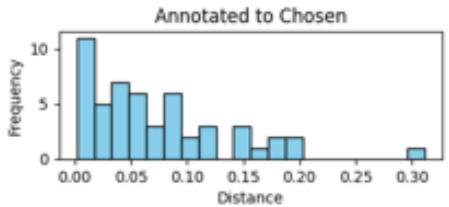
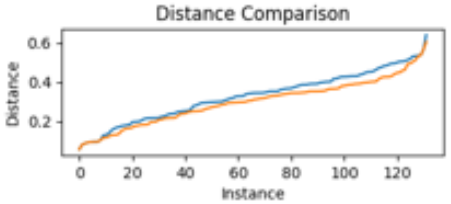
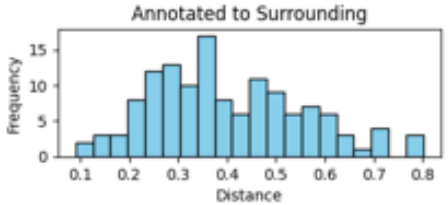
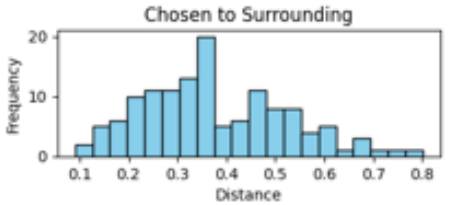
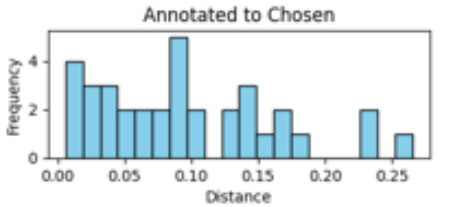
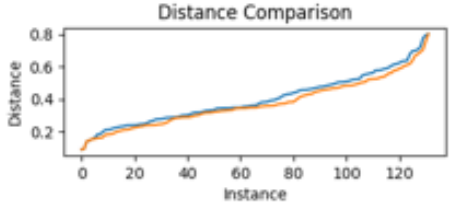
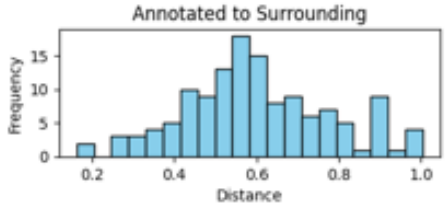
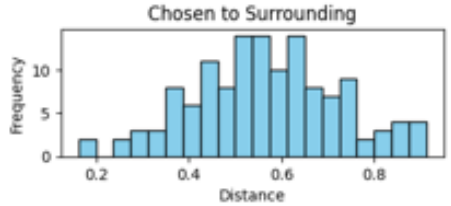
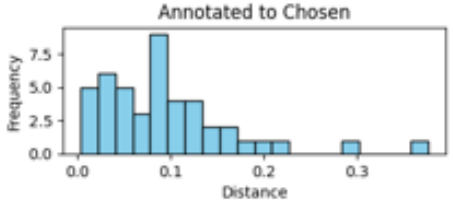
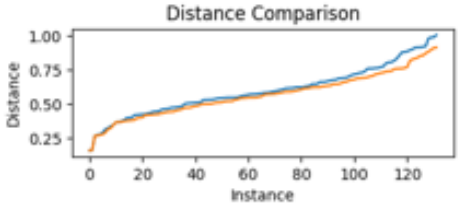
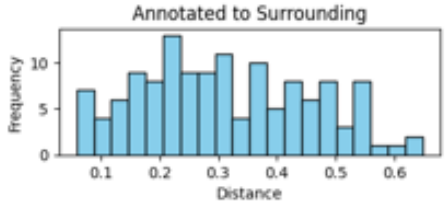
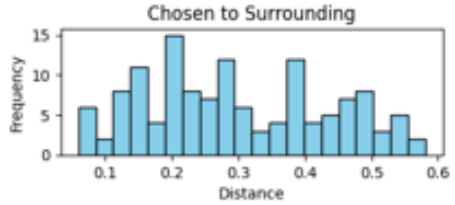
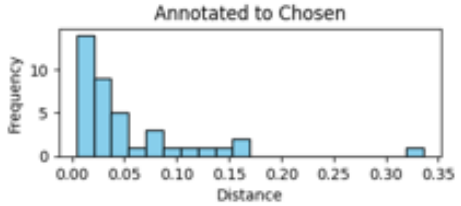
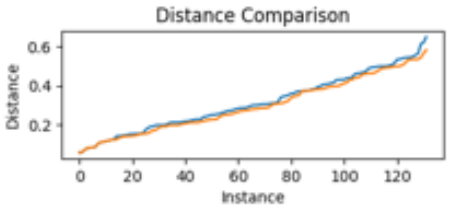
Výsledok choice checkera V2


```
"5C/265/2015__2016-06-02__ECLI:SK:OSNO:2016:5815205480.3": [
    {
        "start": 5085,
        "end": 5100,
        "checker_url": "/SK/ZZ/2005/36/20160101.html#paragraf-78.odsek-2",
        "surrounding_text": "nosti  rodičov  k  deťom je ich zákonná  povinnosť, ktorá  trvá dočasu, kým  deti nie sú sc
        "existing_options": {
            "94/1963": {
                "zakonnik": "94/1963",
                "distance_from_surrounding_text": "0.64754695",
                "version": "20040101.html",
                "provision_text": "94/1963, Funkcia opatrovníka je čestnou povinnosťou."
            },
            "36/2005": {
                "zakonnik": "36/2005",
                "distance_from_surrounding_text": "0.15579677",
                "version": "20160101.html",
                "provision_text": "36/2005, o rodine a o zmene a doplnení niektorých zákonov, výživné, spoločné ustanover
            }
        }
    }
],
```

Výsledky

Model	Total	Matched	Unmatched	Distance	Window size	Accuracy	Avg_Ann_to_Surr	Med_Ann_to_Surr	Avg_Chos_to_Surr	Med_Chos_to_Surr	Avg_Ann_to_Chos	Med_Ann_to_Chos
all-MiniLM-L6-v2	132	80	52	cosine	800	60,60606061	0,327162382	0,34336418	0,298077145	0,308030395	0,073831754	0,06124678
paraphrase-multilingual-MiniLM-L12-v2	132	97	35	cosine	800	73,48484848	0,399497231	0,361825885	0,373926797	0,3483851	0,096437067	0,0884207
distiluse-base-multilingual-cased-v2	132	87	45	cosine	800	65,90909091	0,597916833	0,580486025	0,565329815	0,56241157	0,095588588	0,08853113
kinit/slovakbert-sts-stsb	132	93	39	cosine	800	70,45454545	0,312877411	0,299257395	0,297329678	0,27970916	0,052623097	0,02573776

s_to r	Med_Chos_to_Surr	Avg_Ann_to_Chos	Med_Ann_to_Chos	plot: Annotated to Surrounding	plot: Chosen to Surrounding	plot: Annotated to Chosen	plot: Distance Comparison (orage=chosen:surr, blue=annot:surr)
7145	0,308030395	0,073831754	0,06124678				
5797	0,3483851	0,096437067	0,0884207				
9815	0,56241157	0,095588588	0,08853113				
9678	0,27970916	0,052623097	0,02573776				

Momentálny stav:

- Pozastavená práca na choice checkeri
  - Možnosť pridania vektoru typického použitia
  - Možnosť pridania okolitých zákonov
- Pôvodnú anotovanú sadu, ktorá sa skladala zo 41 rozhodnutí s 132 konfliktnými paragrafmi spomedzi prvých 3500 rozhodnutí sme rozšírili o ďalších 29 rozhodnutí s 95 konfliktnými paragrafmi (spomedzi ďalších 1000 rozhodnutí)
- Aktuálne sa venujeme extrakcii odkazov pomocou promptingu na Európsky súdny dvor, Európsky súd pre ľudské práva, Úrad na ochranu osobných údajov Slovenskej republiky

Aktuálna práca

```
{'7510/04', '3548/04', '11867/09', '18968/07', '30189/07', '28342/95', '42872/02', '9.6.2015', '25132/13',
'7711/06', '33809/08', '36997/05', '72493/10', '46092/06', '2202/05'}

{'Európskym súdom pre ľudské práva', 'Európskom súde pre ľudské práva (ďalej len ESĽP)', 'ESĽP', 'ESĽP '}

{'Tishkevich proti Rusku',
'Tripon proti Rumunsku',
'Kontrová v. SR',
'V.C. v. SR',
'PSMA, s.r.o. proti Slovenskej republike',
'Soltész v. SR',
'Sutyazhnik proti Rusku',
'Abdullayev proti Rusku',
'Luchkina proti Rusku',
'Brumarescu proti Rumunsku',
'Ištván a Ištvánová v. SR',
'COMPCAR, s.r.o. proti Slovenskej republike',
'Labsi v. SR',
'DRAFT-OVA, a.s. proti Slovenskej republike',
'Cornif proti Rumunsku',
'Kutepov a Anikeyenko proti Rusku',
'Černák v. SR',
'Roseltrans proti Rusku',
'Žúbor v. SR',
'DRAFT-OVA, a.s., v. Slovenská republika',
'Kormoš v. SR'}
```

```
Annotated:
rozsudok Súdneho dvora EÚ zo dňa 6. októbra 1982 vo veci CILFIT, C-283/81
uznesenie Súdneho dvora EÚ zo dňa 16. novembra 2010, C-76/10
```

ChatGPT-5

```
{
  "odkazy": [
    {"odkaz":"rozsudok Súdneho dvora EÚ zo dňa 6. októbra 1982, CILFIT, C-283/81, body 9 a 10"},
    {"odkaz":"uznesenie Súdneho dvora EÚ zo dňa 16. novembra 2010, C-76/10, bod 79"}
  ]
}
```

Llama 3.3

```
{
  "odkazy": [
    {
      "text_odkazu": "§ 3 ods. 9 zák. č. 757/2004 Z.z. o súdoch, v znení účinnom od 1.5.2011"
    }
  ]
}
```

Llama 4

```
{
    "odkazy": [
        {
            "text_odkazu": "Nariadenie (ES) č. 805/2004 - Exekučné konanie v Európskej únii"
        }
    ]
}
```

Annotated:
rozsudok SD EÚ zo dňa 5.11.2019, Komisia/Poľsko (Nezávislosť všeobecných súdov), C-192/18, EU:C:2019:924
rozsudky SD EÚ zo dňa 25.7.2018, Minister for Justice and Equality (Nedostatky súdneho systému), C-216/18 PPU
zo dňa 24.6.2019, Komisia/Poľsko (Nezávislosť Najvyššieho súdu), C-619/18
rozsudok SD EÚ z 15. januára 2013,C-416/10, Križan a iní proti Slovenskej inšpekcii životného prostredia
rozsudok SD EÚ zo 6. októbra 2021, C-561/19, Consorzio Italian Management e Catania Multiservizi a Catania Multiservizi
Súdny dvor zaoberal práve v rozsudku C-585/18, C-624/18 a C-625/18
Súdny dvor aj v rozsudku C-824/18
rozsudok C-585/18, C-624/18 a C-625/18
rozsudku C-824/18
rozsudok C-561/19

ChatGPT-5


```
{
  "odkazy": [
    {
      "odkaz": "rozsudok SD EÚ zo dňa 5.11.2019, Komisia/Poľsko (Nezávislosť všeobecných súdov), C-192/18, EU:C:2019:924, bod 101 a citovaná judikatúra"
    },
    {
      "odkaz": "rozsudky SD EÚ zo dňa 25.7.2018, Minister for Justice and Equality (Nedostatky súdneho systému), C-216/18 PPU, EU:C:2018:586, bod 63 a citovaná judikatúra, ako aj zo dňa
                24.6.2019, Komisia/Poľsko (Nezávislosť Najvyššieho súdu), C-619/18, EU:C:2019:531, bod 121 rozsudku"
    },
    {
      "odkaz": "rozsudok SD EÚ z 15. januára 2013,C-416/10, Križan a iní proti Slovenskej inšpekcii životného prostredia, ods. 72"
    },
    {
      "odkaz": "rozsudok SD EÚ zo 6. októbra 2021, C-561/19, Consorzio Italian Management e Catania Multiservizi a Catania Multiservizi, bod 33"
    }
  ]
}
```


Llama 3.3


```
{
  "odkazy": [
    {
      "text_odkazu": "rozsudku C-585/18, C-624/18 a C-625/18"
    },
    {
      "text_odkazu": "rozsudku C-824/18"
    }
  ]
}
```


Llama 4


```
{
    "odkazy": [
        {
            "text_odkazu": "rozhodnutie Najvyššieho správneho súdu SR"
        }
    ]
}
```

Anotovanie súdnych rozhodnutí

mimoriadneho opravného prostriedku, ktorý nebol dostupný účastníkom konania, ale len generálnemu

prokurátorovi, ktorý mal, čo sa týka aplikácie tohto inštitútu voľnú úvahu (rozsudok ESĽP vo veci Brumarescu

reference office sides

proti Rumunsku, sťažnosť č. 28342/95, rozsudok zo dňa 28. októbra 1999). Z judikatúry ESĽP ďalej vyplýva, že

id date

prieskum právoplatných súdnych rozhodnutí musí rešpektovať princíp právnej istoty, keď konečné rozsudky by

mali vo všeobecnosti zostať nedotknuté, zrušené môžu byť iba za účelom nápravy fundamentálnych chýb

(rozsudok ESĽP vo veci Tishkevich proti Rusku, sťažnosť č. 2202/05, rozsudok zo dňa 4. decembra 2008).

reference office sides id date

Fundamentálne pochybenie, ktoré opodstatňuje zrušenie rozhodnutia, môže podľa názoru ESĽP predstavovať

napríklad omyl v právomoci, vážne porušenie súdneho procesu alebo zneužitie právomoci (rozsudok vo veci

reference

Luchkina proti Rusku, sťažnosť č. 3548/04, rozsudok zo dňa 10. apríla 2008). Mimoriadny prieskum by ale

sides id date

Aktuálna práca

- Zväčšenie anotovanej sady
- Osvojenie si praktík prompt engineeringu
- Testovanie rôznych modelov a promptov
- Rozparsovanie extrahovaného výsledku z modelu na jednotlivé entity (offsets, date, title, office, ID)

Literatúra

1. Dávid Varga, et al. „Extraction of Legal References from Court Decisions“. Vol. 3498, 2023, pp. 89–95,

2. Darji Harshil, Jelena Mitrović, and Michael Granitzer. "A Dataset of German Legal Reference Annotations." Proceedings of the Nineteenth International Conference on Artificial Intelligence and Law. 2023.

3. Elena Leitner, Georg Rehm and Julian Moreno-Schneider. “A Dataset of German Legal Documents for Named Entity Recognition“. Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)

4.Malte Ostendorff, Till Blume, and Saskia Ostendorff. 2020. „Towards an open platform for legal information“. In Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020. 385–388.