

Neviditelné a užitečné: Datová problematika právních informačních systémů

Tereza Novotná

Osnova

1. Motivace
2. Právní informace a data – legislativa
3. Právní informace a data – judikatura
4. Právní informace jako základ pro PIS
 1. Korpusy či datasety právních informací
 2. NLP metody zpracování dat
 3. Natrénování modelu
 4. Připojení uživatelského rozhraní
5. Příklady dobré praxe ze zahraničí
6. Právní informace dostupné v ČR – korpusy a datasety
7. Use case – příklad ChatGPT a řešení právní otázky

Motivace

- Důležitost právních dat a informací - *conditio sine qua non* pro jakýkoliv právní informační systém
- *Z právního hlediska*: povinnost seznámit se s právními pravidly a následovat je (neznalost zákona neomlouvá) vs. nesnadný přístup k právním pramenům a velké množství pramenů
- *Řešení*: Dostupné pokročilé metody (založené nejen na AI) s potenciálem zlepšit přístup a vyhledávání (právních) informací

Právní informace a data - legislativa

– Data vs. Informace

– Co jsou?

- Primární, sekundární prameny - jejich TEXT
- Data a metadata (objektivní, subjektivní)

– Kde se berou?

- Legislativní proces ústí do publikačního procesu
- Promulgace práva v ČR: [zákon o sbírce zákonů a mezinárodních smluv č. 309/1999 Sb.](#), promulgace skrz fyzickou sbírku zákonů, k nahlédnutí na úřadech
- [Nový zákon o sbírce zákonů a mezinárodních smluv č. 222/2016 Sb.](#) – několikrát odložená účinnost, počítá s elektronickou sbírkou

Právní informace a data - judikatura

- Publikace judikatury na stránkách soudů ([NS](#), [NSS](#), [ÚS](#))
- Publikace judikatury nižších soudů online – povinnost od července minulého roku na [justice.cz](#)
- Webové databáze neúplné, nedokonalé vyhledávací možnosti
- Sbírky rozhodnutí NS, NSS a ÚS
- Proč je důležitý hromadný přístup k právním datům?

Právní informace jako základ pro PIS

- Informace musí být v systému *obsažena* vs. informace musí být v systému *vyhledatelná* (otázka existence informace vs. otázka zpracování informací pro vyhledávání)
- *Relevance* vyhledávaných informací
- *Intuitivnost* vyhledávacích metod a uživatelského rozhraní

Právní informace jako základ PIS

- Pokročilé metody zpracování textu – *natural language processing* (NLP) jsou **závislé na dostupných textových datech**
- NLP metody jsou nejčastěji založené na strojovém učení, které je přímo závislé na **velikosti a kvalitě** textových dat

Právní informace jako základ PIS

1. Korpusy či datasety právních informací (velikost, kvalita, úplnost atd.)
2. NLP metoda zpracování dat
3. Natrénování modelu
4. Připojení uživatelského rozhraní

1. Korpusy či datasety právních informací

- Data
 - strukturovaná, semistrukturovaná, nestrukturovaná
 - Anotovaná, neanotovaná, anonymizovaná
 - Text (celý dokument), segmenty, metadata
- Datové formáty
 - XML
 - CSV
 - TXT
- Dostupnost a přístup k datům
 - API
 - Hromadné stažení
 - RSS kanály
- Problém s obecnou nedostupností právních dat u nás

2. NLP metoda zpracování dat

- Cíl zpracování textu: kategorizace, tematické členění, vyhledávání, návrhy podobných dokumentů, sumarizace, generování textu
- Dělení metod: supervidované, semisupervidované, nesupervidované
- Otázka manuálního zpracování či manuální evaluace (proto je ChatGPT o tolik lepší než předchozí modely)

3. Natrénování modelu

- Předzpracování textu (tokenizace, tagování, anotování manuální či automatické, určení slovních druhů, vyjmutí určitých slov atd.)
- Samotné trénování modelu – dle metody, dnes se používají nejčastěji metody založené na pravděpodobnosti výskytu slov vedle sebe v textu
- Definování parametrů, trénování

4. Uživatelské rozhraní

- Intuitivnost
- Interakce s uživatelem
- Sběr zpětných dat a jejich užití pro zpřesnění metody a přetrénování modelu (evaluace, uživatelská data, dotazníky atd.)
 - Cena za službu „zdarma“

Příklady z (různě dobré) praxe

- [CanLII](#)
- [Légifrance](#)
- [Harvey](#)
- [LexisNexis](#)

Datasey a modely dostupné u nás

- [Czech Court Decisions Corpus](#)
- [Czech Court Citations Dataset](#)
- Manuálně anotované datasey pro segmentaci a extrakci citací:
 - [Annotated Corpus of Czech Case Law for Reference Recognition Tasks](#)
 - [Annotated Corpus of Czech Case Law for Segmentation Tasks](#)
- LDA a NMF model pro tematické modelování českých soudních rozhodnutí ([github stránka](#))

Use case - ChatGPT

- Generativní model
- Vycházející z dat, která mu byla poskytnuta při trénování
- Problematické odpovídání na konkrétní právní otázky
- Obecnější problém právního jazyka:
 - Dostupnost datasetů
 - Interpretace textu
 - Sémantika textu
 - Synonyma

Use case - ChatGPT



Jak je v českém právním systému upraveno zvyšování nájemného bytu?



V českém právním systému je zvyšování nájemného bytu upraveno zákonem o nájemném a bytech (č. 89/2012 Sb.).



Podle tohoto zákona má nájemce právo užívat byt za sjednanou nájemní cenu a nájemce je povinen platit nájemné ve výši stanovené nájemní smlouvou. Pokud nájemce plní své povinnosti, pronajímatel nemůže nájemné zvýšit.

Pokud pronajímatel chce nájemné zvýšit, musí nájemce o této skutečnosti informovat písemně s výpovědní lhůtou minimálně šest měsíců před plánovaným zvýšením. Zvýšení nájemného pak musí být řádně odůvodněno, musí odpovídat tržní ceně a nesmí být neúměrné.

Nájemce má poté právo na přezkoumání zvýšení nájemného u soudního nebo správního orgánu. Pokud soud nebo správní orgán shledají, že zvýšení nájemného bylo neúměrné, může rozhodnout o jeho snížení.

Je důležité dodat, že nájemce může souhlasit se zvýšením nájemného dobrovolně a bez výpovědi smlouvy, avšak pouze pokud jsou splněny podmínky uvedené v zákoně.

Děkuji za pozornost.

Budete-li mít nějakou otázku, obraťte se na mě prosím na: tereza.novotna@law.muni.cz a nebo přijďte na konzultační hodiny!